# Combining Evolutionary Algorithms and Agent-Based Simulation for the Development of Urbanisation Policies

Marta Vallejo

submitted for the degree of
Doctor of Philosophy

Heriot-Watt University
School of Mathematical and Computer Sciences

July 2017

# Abstract

Urban-planning authorities continually face the problem of optimising the allocation of green space over time in developing urban environments. To help in these decision-making processes, this thesis provides an empirical study of using evolutionary approaches to solve sequential decision making problems under uncertainty in stochastic environments. To achieve this goal, this work is underpinned by developing a theoretical framework based on the economic model of Alonso and the associated methodology for modelling spatial and temporal urban growth, in order to better understand the complexity inherent in this kind of system and to generate and improve relevant knowledge for the urban planning community. The model was hybridised with cellular automata and agent-based model and extended to encompass green space planning based on urban cost and satisfaction.

Monte Carlo sampling techniques and the use of the urban model as a surrogate tool were the two main elements investigated and applied to overcome the noise and uncertainty derived from dealing with future trends and expectations. Once the evolutionary algorithms were equipped with these mechanisms, the problem under consideration was defined and characterised as a type of adaptive submodular. Afterwards, the performance of a non-adaptive evolutionary approach with a random search and a very smart greedy algorithm was compared and in which way the complexity that is linked with the configuration of the problem modifies the performance of both algorithms was analysed. Later on, the application of very distinct frameworks incorporating evolutionary algorithm approaches for this problem was explored: (i) an 'offline' approach, in which a candidate solution encodes a complete set of decisions, which is then evaluated by full simulation, and (ii) an 'online' approach which involves a sequential series of optimizations, each making only a single decision, and starting its simulations from the endpoint of the previous run.

# Acknowledgements

I would like to thank everybody who contributed in any positive way to the creation of this thesis and the related research. To my main supervisor Prof. David W. Corne for giving me the freedom to conduct my PhD in my personal way and provided me special support and useful feedback at the end of the writing period.

To my second supervisor Prof. Verena Rieser who contributed specially in the first years of this project both, directly to this research and indirectly in other extra-curricula activities for her constant enthusiasm and ideas.

To other people who read some of my papers and/or thesis like Dr. Michael Lones, Dr. Matthias Keysermann, and Alex Duff. I am very grateful to all of you. I would like to say thank you to my examiners Dr. Julia Handl and Prof. Greg Michaelson for all of their guidance through this process. Thanks to them this thesis was substantially improved and refined.

I would especially like to thank to Alistair McConnell for the love, support, and constant encouragement he gave me during the last step of this long project. Finally I would like to mention the unconditional support that I have always received from my mother Cristina de Frutos during these years and beyond and dedicate this thesis to her.

ACADEMIC REGISTRY
**Research Thesis Submission**

| | | | |
|---|---|---|---|
| Name: | *Marta Vallejo* | | |
| School/PGI: | MACS/School of Mathematical and Computer Sciences | | |
| Version: *(i.e. First, Resubmission, Final)* | Final | Degree Sought (Award **and** Subject area) | *Philosophy of doctore degree* |

## Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

| | | | |
|---|---|---|---|
| Signature of Candidate: | | Date: | 11/07/2017 |

## Submission

| | |
|---|---|
| Submitted By *(name in capitals)*: | Marta Vallejo |
| Signature of Individual Submitting: | |
| Date Submitted: | |

## For Completion in the Student Service Centre (SSC)

| | | | |
|---|---|---|---|
| Received in the SSC by *(name in capitals)*: | | | |
| *Method of Submission* *(Handed in to SSC; posted through internal/external mail)*: | | | |
| *E-thesis Submitted* (***mandatory for final theses)*** | | | |
| Signature: | | Date: | |

Please note this form should bound into the submitted thesis.

Updated February 2008, November 2008, February 2009, January 2011

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

**$\mu$GA** Micro-Genetic Algorithm.

**ABM** Agent-Based Model.

**AHP** Analytical Hierarchy Process.

**ANN** Artificial Neural Network.

**CA** Cellular Automata.

**CBD** Central Business District.

**CDM** Cognitive Driver Model.

**DEC-POMDP** Decentralized Partially Observable Markov Decision Process.

**DEED** Dynamic Ecological Exurban Development.

**DP** Dynamical Planning.

**DT** Decision Tree.

**DUEM** Dynamical Urban Evolutionary Modelling.

**EA** Evolutionary Algorithm.

**EP** Evolutionary Programming.

**ES** Evolutionary Strategy.

**FL** Fuzzy Logic.

**GA** Genetic Algorithms.

**GIS** Geographical Information Systems.

**GP** Genetic Programming.

**HFLP** Hierarchical Facility Location Problem.

**ID** Influence Diagram.

**ILUTE** Integrated Land Use, Transportation, Environment.

**LA** Location-Allocation.

**LCD** Landcover Deltatron Model.

**LP** Linear Programming.

**LSUM** Large Scale Urban Model.

**MCE** Multi-criteria Evaluation.

**MCP** Maximum-k Covering Problem.

**MDP** Markov Decision Process.

**MEM** Matter-Element Model.

**MOEA** Multi-Objective Evolutionary Algorithm.

**MOGA** Multiple Objective Genetic Algorithm.

**MSD** Maximum Service Distance.

**MUCM** Multilayer Urban Canocopy Model.

**NPGA** Niched Pareto Genetic Algorithm.

**NSGA** Non-dominated Sorting Genetic Algorithm.

**PAES** Pareto-archived Evolutionary Algorithm.

**POMDP** Partially Observable Markov Decision Process.

**RBFNN** Radial Basis Function Neural Network Model.

**RDGA** Rank-Density based Genetic Algorithm.

**RIKS** Regional Integration Knowledge System.

**RS** Repast Simphony.

**SA** Simulated Annealing.

**SAMBI** South Atlantic Migratory Bird Initiative.

**SDMP** Sequential Decision Making Problem.

**SPEA** Strength Pareto Evolutionary Algorithm.

**SPEA2** Strength Pareto Evolutionary Algorithm II.

**TS** Tabu Search.

**UGM** Urban Growth Model.

**UHI** Urban Heat Island.

**WULUM** Water-Use and Land-Use Model.

# Chapter 1

# Introduction

## 1.1 Overview

Green areas can be seen as green lungs that significantly contribute to a varied range of social, economical and environmental aspects in densely populated areas (Chiesura, 2004; Bowler et al., 2010a). Numerous studies highlight how human interactions with nature are beneficial for physical, social, and mental wellbeing (Sop Shin, 2007; Takayama et al., 2014), their crucial impact on the economy, quality of life and in the local climate of the cities (Costanza et al., 1998; Nowak and McPherson, 1993) among others. Currently the necessity of open spaces is recognised more intensively by the society due to the appearance of new social values and the rising standard of living (del Saz Salazar and García, 2007). Due to that, the study of mechanisms that can mitigate the ecological degradation that is invariably linked with urban expansion is one of the most urgent research issues within the broad field of urban planning.

## 1.2 Motivation

Pressure on green spaces close to or within city boundaries is significant and likely to grow (Glickman, 1999). These areas have been increasingly recognised as important components of urban ecosystems, providing various kind of important environmental and social services (Haviland-Jones et al., 2005; Takayama et al., 2014) as was reviewed previously. Hence, it is vital for planners and decision-makers that the provision of these areas is performed maximising the impact, the benefits and the attractiveness of each of the selected parcels (Uy and Nakagoshi, 2008) in compliance with a sustainable urban development.

However, to fully understand the interactions of the involved complex phenomena and be capable of dealing with a large number of environmental and socio-economical constraints, scientists need a better and larger set of ecologically meaningful methods that can be applied to spatial multi-criteria evaluation and conservation decision making. In this regard, the use of models and interactive computer-based systems (Church, 2002) can be mentioned. These models could both explore and extrapolate the dynamics of the system to infer future trends or also understand the nature of the processes within it.

## 1.3 Research aims & Objectives

The primary objective of this study is to evaluate how evolutionary algorithms can be applied to dynamical location-allocation problems under uncertainty, with a special focus on urban green spaces. This question is intended to be addressed from a spatial and temporal dimension to be able to help in planning decision-making processes. This research is performed primarily from a methodological and theoretical perspective using simulation techniques and different types of planning procedures.

A solution to the problem analysed in this thesis yields a useful tool to urban planners and decision-makers, who can first study and develop efficient conservation designs that can capture the nature of the process involved in their own complex decision-making objectives. After the construction of that schedule with a suitable list of land purchases, the policy can be executed to obtain the most cost-effective and efficient implementation of that design.

This dissertation will be guided by the following central questions:

- Can the evolutionary algorithm be defined to be able to deal with the uncertainty derived from creating urban plans where future conditions are taken into account?

- Can the evolutionary algorithm be configured to outperform intelligent greedy algorithms that take advantage of the adaptive submodularity that correspond with some of the configuration of the problem under consideration?

- Could we use different planning approaches, where the decisions are taken at a time and beforehand to come up with accurate solutions?

## 1.4 Contributions

This multidisciplinary thesis has led to contributions in both, computational intelligence and computational sustainability fields. These contributions can be summarised as follows:

### 1.4.1 Computational Intelligence

- In this thesis, new algorithms are designed for the task of sequential decision making under uncertainty in complex, stochastic environments, where the effects of earlier decisions have a highly non-linear relationship with the value of later decisions. More specifically, an 'offline' algorithm and an 'online' algorithm are designed and evaluated for interconnected sequential location-allocation problems. These algorithms are applied to a case study in urban development, which is underpinned by a complex simulation environment full of inter-connected uncertainties.

  Each of these algorithms hybridises evolutionary search with a Monte Carlo sampling strategy in order to mitigate the uncertainty involved in the system. This information is required for the potential effects of each decision and their fitness evaluation.

  - The offline algorithm addresses the planning problem by searching the space of complete decision sequences. It is shown that the approach, more traditional in style, can be applied successfully to such type of problems as long as a suitable sampling strategy is involved to provide necessary structure to the search space.

  - In contrast, the 'online' algorithm involves a sequence of optimisation runs; each run in the sequence searches the space of 'next' decisions, informed by the decisions that have been fixed by earlier runs. The online algorithm is more unusual in style, and (on the case study examined) is significantly faster without significant loss of solution quality. The online algorithm therefore brings us closer to an approach that would be usable by practitioners in cases where such tools had not been previously available due to the time complexity of suitably sophisticated simulations.

- The characterisation of the 'offline' algorithm as *Adaptive Submodular* and the consequences of this property for the performance of different types of optimisation strategies are analysed. The conclusions show that this effect is linked to the level of complexity of the topological configuration of the city along with the level of uncertainty of the problem. The phenomenon can shed light into the problem of creating effective plans under uncertainty for the urban community.

### 1.4.2 Computational Sustainability

- A related contribution is the urban growth model itself. The model described and developed in this thesis is a theoretical variant of a reduced open version of the published in Murray-Rust et al. (2013), and it has been extended primarily with new components and rules to reflect the value (and dynamics of that value) of green spaces in the context of the urban spatial structure theory of Alonso (1964). These extensions add to those other models, where Alonso's theory has been incorporated, knowledge particularly in terms of ecological valuation and the effect of connectivity between green areas.

- The solution to the urban planning problem analysed in this thesis would yield a useful tool for urban planners and decision-makers. One of the problem faced by such professionals involves the design of suitable policies for green-space allocation that will lead, in each of the short, medium and longer terms, to urban plans that, while growing and changing, remain effective and positive for all stakeholders. Simulation tools in themselves provide helpful instruments for analysing the effect of hand-designed policies; however, optimisation techniques wrapped around such tools have the potential for much greater contributions to the planning world. In this thesis, it is contributed two such algorithms, which we have shown to be effective (specially the online one) for discovering good-quality sequences of decisions in reasonable time for a complex urban model. Meanwhile, the findings of experiments with the online and offline algorithms reveal that the algorithms use interestingly different strategies, and this in turn reflects the usefulness of such a model as a discovery engine for urban planners. Finally, our results show that, to the extent that Alonso's model (and the implementation in this thesis) is valid, the spatial clustering of parks tends to correlate with higher levels of satisfaction from visitors, in contrast with the satisfaction levels reported for a large number of scattered green areas. This finding is at odds with received wisdom in the urban planning world, and is potentially a valuable contribution to debate in that area.

## 1.5 Outline of the Thesis

The following list provides the organisational structure for the remaining chapters of the present thesis, which covers:

1. Chapter 2 provides the theoretical background information regarding the major conceptual elements that the thesis uses in the different parts that this research presents. This conceptual outline includes a description of urban models, types and evolution. This is followed by the definition of Alonso's economic urban model and its related *centricity* concept. Afterwards, the use of a Cellular Automata as a tool to represent spatio-temporal dynamics along with Agent-Based Systems, which are able to depict population dynamics based on heterogeneous individual choices are presented. The conceptualisation of two

types of problems: the Location Allocation problem and a dynamical variant, that can be considered a Sequential Decision Making problem, are introduced. Finally, the general concepts about Evolutionary Algorithms in its single and multi-objective optimisation version with special attention to their previous use in urban studies are summarised.

2. Chapter 3 depicts the theoretical urban growth model used in this thesis as a tool to understand different problems by using simulation techniques. The chapter includes information about its different elements named population, urban, rural and ecological dynamics and how the interactions between them create different types of non-linear relationships that will critically affect the different optimisation processes.

3. Chapter 4 illustrates multiple variants of noise and uncertainty that evolutionary algorithms can face and different tools to overcome them. An especial focus on fitness approximation and Monte Carlo sampling is done since these techniques are used in the algorithms described in this thesis. The data gathering and the statistical data generation are depicted in detail, along with the different sources of uncertainty covered in the theoretical model. Finally the way the techniques are applied and the statistical significance of the generated results are also described.

4. Chapter 5 is aimed at the description of the methodology used to support the spatial allocation planning of green spaces in an urban area, using a non-adaptive version of an evolutionary strategy for optimisation purposes under uncertainty. The chapter starts with the formal description of the problem and of its submodular nature. It continues with the analysis of the consequences for the optimisation techniques that this adaptive submodularity property implies. Specially how an offline evolutionary algorithm may be outperformed by simple intelligent greedy algorithms in some simple versions of the problem is shown and explained. Finally, different scenarios are selected to test multiple levels of complexity in relation with different parameters of the model. The generated

results serve to further understand how the different optimisation techniques behaves under different topological circumstances and levels of uncertainty.

5. Chapter 6 investigates how evolutionary algorithms can also be implemented using, in this case, an online perspective to search for more efficient urban planning strategies. The online approach has to tackle the problem from a different perspective, having to deal with less level of uncertainty, with additional constraints in terms of computational time used in each generation, can be used to analyse the problem from a different perspective. The comparison of the results from using the offline version help us to further understand the problem, highlighting the role of the budget that it is provided each turn by the corresponding governmental entity and how the optimisation process is strongly dependent on the saving of financial resources. This budget, that it is received periodically, has an accumulative nature since the amount that it is not used when it is received, is added to the funds available for land purchasing purposes. The discovery of the most convenient moments in which these funds should be used could be a major additional factor to study as well as the acquisition of patches of land.

6. Chapter 7 concludes with a list of the main findings and the inclusion of the most important extensions and refinements that the model and the optimisation procedure could undertake. From the sustainability point of view, the aspects that may prevent visitors going to a given park and the enrichment of the agents profile to study different usability profiles will be discussed. On the other hand the computational intelligent aspects of the present thesis could be extended with the development of the multi-objective optimisation version of the algorithm, with the concrete focus on the development of a new fitness function aimed at preserving the patches of land with more ecological value in the area under consideration.

## 1.6 Publications Resulting From this Research

Vallejo, M. and Corne, D. (December 2016) "Evolutionary Algorithms under Noise and Uncertainty: a location-allocation case study", *The 2016 IEEE Symposium Series on Computational Intelligence (IEEE SSCI) under the 2016 IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (IEEE CIDUE'2016).* Athens, Greece.

Vallejo, M. and Corne, D and Vargas, P. (December 2016) "Online/Offline Evolutionary Algorithms for dynamic urban green space allocation problems", *Journal of Experimental & Theoretical Artificial Intelligence (JETAI).*

Vallejo, M. and Rieser, V. and Corne, D. (June 2015) "Genetic Algorithm evaluation of green search allocation policies in multilevel complex urban scenarios", *15th International Conference on Computational Science (ICCS'15) Reykjavík, Iceland. Selected for a special issue.* Journal of Computational Science (JOCS), Volume 9, pp 57-63. Elsevier

Vallejo, M. and Rieser, V. and Corne, D. (January 2015) "Agent-Based Modelling for Green Space Allocation in Urban Areas: Factors Influencing Agent Behaviour", *7th International Conference on Agents and Artificial Intelligence (ICAART).* Lisbon, Portugal.

Vallejo, M. and Rieser, V. and Corne, D. (October 2014) "Evolving optimal spatial allocation policies for complex and uncertain environments", *Agents and Artificial Intelligence. In: Communications in Computer and Information Science (CCIS)*, Volume 449, pp 351-369. Springer-Verlag.

Vallejo, M. and Corne, D. and Rieser, V. (February 2013) "Evolving Urbanisation Policies: Using a Statistical Model to Accelerate Optimisation over Agent-Based

Simulations", $5^{th}$ *International Conference on Agents and Artificial Intelligence (ICAART)*. Barcelona, Spain.

# Chapter 2

# Background

## 2.1 Introduction

In this chapter we introduce various background theory and concepts which underpin the research in the remainder of the present thesis. The first section presents a description of green spaces and their related benefits along with the general concept of modelling as a widely used tool in the context of urban systems. Here, the related theory about urban models is discussed, including the conception of a city as a complex system, and a discussion of the most notable types of urbanisation models and their evolution. Key elements that will be introduced are the macroeconomical model of Alonso, and the concept and consequences of the 'centricity' assumption. The usability and scope of theoretical models will be also covered.

The second section then focuses on computational implementations of the background theory. Concretely, we introduce and formulate a framework comprising Cellular Automata and Agent-Based Systems, which is typically applied in the context of urban development to study the spatial and temporal dynamics and patterns of land-use, land inhabitants, and their inter-relationship. The chapter continues with a description of the two major optimisation tasks investigated in this document: (i) the location-allocation problem, which is focused on the optimisation of the placement of certain facilities according to a set of pre-defined goals, and (ii) the sequential decision-making problem under uncertainty, which captures the notion of determining a planning policy for a given time horizon, despite significant uncertainty about how the environment will develop over time.

To conclude the chapter, we introduce and discuss Evolutionary Algorithms and their applications in general terms, covering both single-objective and multi-objective optimisation scenarios, focusing specifically on the application of the Pareto-archived Evolutionary Algorithm (PAES) as an algorithm to solve multi-objective problems in an urban context.

## 2.2 Green Areas

Natural landscape covered with natural or man-made vegetation inside or close to urban zones covers concepts like county parks, areas of outstanding natural beauty, natural reserves, forest parks or crown lands (Allison, 1975). However, not all natural landscapes located close to urban or peri-urban areas are protected with the only purpose of maintaining the biodiversity in a large scale. There are other types of land configuration which can be included into the previous list contributing at small scale. This areas include street trees, lawns/parks, urban forests, cultivated land, wetland, lake shore/seashore, streams and golf courses (Bolund and Hunhammar, 1999). These small patches, normally isolated from each other by a urban matrix, are considered as an independent ecosystem called a city ecosystem (Ahern, 1995).

## 2.3 Benefits of Green Spaces

The conservation and the promotion of green spaces that are located close to or within urban areas cannot be fully understood without summarising the amount of benefits derived from them. This wide range of advantages can be grouped according to different type of services they generate. In this regard, these services can be classified into three sections namely ecosystem services, social and recreational factors and economic benefits. An explanation of each type is summarised in the following list:

### 2.3.1 Ecosystem services

- Urban areas are considered as one of the main sources of pollution. Green areas serve as a pollution control by affecting gas cycles (Lebel et al., 2007; McHale et al., 2007), by absorbing part of green-house gasses and by fostering carbon sequestration (Balvanera et al., 2005; Niemelä et al., 2010). These positive effects are mainly caused by the presence of trees, especially in the case of forests (Bernatzky, 1983). In this regard, an empirical study conducted by the Spanish Environmental Ministry (Ministerio de Medio Ambiente, 2000) shows that one hectare of Mediterranean forest is capable of absorbing 3.7 tonnes of $CO^2$ per year (global emissions from fossil fuel and industry: $36.3 \pm 1.8$ Gigatonnes of $CO_2$ in 2015 (Le Quéré et al., 2015)). Another significant example of this role is the fact that in certain cases green areas can filter more than 80% of air pollutants (Bolund and Hunhammar, 1999; Zhou and Rana, 2012). Hence, green areas and forests can be a real measure against the spread of these noxious gasses and the negative impacts of climate change (Donovan and Butry, 2009; Bowler et al., 2010a).

- Green areas have a potential to regulate air temperatures, creating a cooling effect and combating the Urban Heat Island (UHI) effect at a city level by means of shade, evapotranspiration and with the enhancement of local wind breath (Jo and McPherson, 2001; Doick et al., 2014). This is produced by the beneficial effects spread beyond the boundaries of the green area (Taha et al., 1989; Saito et al., 1991). Another consequence of this effect is the decrement of the risk transmission of infections diseases. When urban temperature growths, this might accelerate the development of some animals, reducing their larval period and the scales at which they will be spread.

- The increment of impermeable surfaces like paved areas can provoke severe problems of flooding. Forest and green plants may regulate rainfall and improve water retention by interception and evapotranspiration processes (Bolund and Hunhammar, 1999; Guo et al., 2000; Nosetto et al., 2012) which can make cities less vulnerable to these natural hazards. Another consequence of the spread of these surfaces is the emission of storm water run-off pollution like nitrates ($NO_3^-$) where green plants can aid to remove or retain them by 30% (Wise, 2008).

- These areas also help maintaining the level of humidity in the atmosphere and restraining soil erosion. Mentionable is also their role in noise abatement by means of reducing the level of noise and mitigating the effect of light pollution which causes disturbances in biota rhythms.

- Wetlands, which are considered a type of green area, support water protection and improve its quality (Smith et al., 2002) by fostering groundwater supply, reducing the runoff, increasing the level of infiltration of the soil, suspension and storage (Brauman et al., 2007).

### 2.3.2 Social and recreational factors

People with different profiles characterised by their gender, age and socio-economic status may differ in how they use and perceive green areas (Burke et al., 2009; Eisler et al., 2003). Consequently, it is a complex task to properly fulfil the diverse range of demands that can arise from them. However, multiple social and recreational benefits that green spaces provide to their potential visitors were highlighted:

- Green areas can improve the quality of life of the visitors by fostering physical and mental fitness including the practise of sports and games (Payne et al., 2005; Bedimo-Rung et al., 2005). Takayama reported higher vigour and subjective improvement of vitality even in short-term walking (Takayama et al., 2014).

- People who visit parks regularly are more likely to improve both perceived and objective general health (Bowler et al., 2010b; Lee and Maheswaran, 2011), specially in young and elderly people (Godbey et al., 1992) and in pregnant women (Dadvand et al., 2012, 2014). Besides, frequent users exhibit longevity (Takano et al., 2002; Maas et al., 2006) even if the relationship between both factors is not completely understood yet (Health Council of the Netherlands and RMNO, 2004). Besides, in the presence of green views, patients tend to improve their recovery process in different ways (Park and Mattson, 2008; Ulrich et al., 1991). Additionally, visiting green areas fosters the immune system (Li et al., 2007a), decreases the blood pressure (Park et al., 2009) and reduces incidence of allergies in children (Hanski et al., 2012). Also these areas can balance socioeconomic health inequalities in deprived population (Mitchell and Popham, 2008; Ward Thompson and Aspinall, 2011).

- From the point of view of the psychological aspects that visitors can get advantage from these facilities it can be mentioned that parks can provide relaxation and relief from stress (Ulrich, 1981; Health Council of the Netherlands and RMNO, 2004; Li et al., 2007b), sense of freedom and refuge, silence, emotional peace, psychological equilibrium (Kaplan, 1985) and positive mood (Haviland-Jones et al., 2005; Pretty et al., 2005; Barton and Pretty, 2010) by means of the contact with and enjoyment of nature (Furlan, 2004). It is also reported the improvement of the level of anxiety and depression in individuals (Maas et al., 2009), facilitating psychological restoration (Bodin and Hartig, 2003; Bowler et al., 2010b). There are studies that go further stating the spiritual component linked with this areas. From this point of departure, green areas can be included into a metaphysical or spiritual dimension (Ward Thompson, 2002), inspire a moral attitude to connect humans with nature and serve as an energy source (Chiesura, 2004).

- Regarding the mental benefits associated with green areas it can be mentioned that they produce attention restoration (Hartig and Staats, 2003), reduce mental fatigue, improve problem-solving skills and concentration (Kaplan, 1990; Herzog et al., 1997). Here, their potential can be also included as a tool for educational and scientific activities and the accumulation of knowledge that these areas can provide (Millenium Ecosystem Assessment, 2003; Bolund and Hunhammar, 1999).

- They are also considered as a place for social encounters, which foster community integration (Coley et al., 1997; Kuo et al., 1998; Leyden, 2003; Kawachi and Berkman, 2000) as a result of the practise of activities like picnicking, dog walking, bird watching, family activities, or photography among others. Reduction in crime rates are also reported (Kuo and Sullivan, 2001) by being responsible for lower anger and aggression scores (Hartig et al., 1991).

- With reference to their aesthetic benefits, green areas provide balance from the visual urban impact, which allows visitors the contemplation of a pleasant atmosphere, enhanced views and scenery. They can be also seen as a source of preservation of scenic views, historical and archaeological sites and rich environmental areas for future generations.

### 2.3.3 Economic benefits

Green areas have a positive impact on the economy in many ways. In this section, the most important factors are summarised:

- Cities with highly valuable recreational areas can be promoted as an attractive tourist destination to visit due to their nature-based attractions (Ceballos Lascurain, 1996; Chiesura, 2004). From a touristic perspective, these areas can generate multiple business opportunities including tour guides services, overnight visits and commercial activities within them like food concessions, kiosks, cafés and so on.

- The beneficial effects over the health of the population may save substantial health costs in the long term (Ward Thompson, 2002).

- The saving in electricity consumption from reducing air conditioning usage and heating energy due to the cooling effect and the trees capacity of reducing wind speed (McPherson and Simpson, 2003) can be also mentioned.

- The increment in productivity and the reduction of level of stress in employees that have their working place equipped with nature views (Kaplan and Kaplan, 1989; Lewis, 1996; Leather et al., 1998; Sop Shin, 2007) were reported.

- Green areas are the source of provision of goods such as timber, food and water (Brauman et al., 2007) and, from this perspective, they also contribute to the economy.

### 2.3.4 Urban Planning

A key element in the study of urban evolutionary patterns is the peri-urban areas. A peri-urban area can be defined as a transition zone where different land types like agricultural and forestry coexist with urban residences, industry, transportation and leisure areas. In the long term, this urban-rural fringe tends to be transformed rapidly into new built-up zones and consequently additional peri-urban areas emerge around them, which provokes a phenomenon called urban sprawl (Mills, 1981) characterised by low-density, dispersed and discontinuous land development in the urban fringe. The continued use of peri-urban transition zones for the construction of new settlements entails the fragmentation and segregation of open areas that, in certain cases, can allocate very valuable resources. This unsustainable process of natural land degradation should be controlled by nature conservation plans and green belt legislations, based on the fact that high speed processes in land-cover change are associated with negative effects and rapid shifts in the original biodiversity (Grimm et al., 2008). One of the facts that makes this control process particularly difficult is that the urban expansion dynamics need to be analysed at many time-scales to be able to accomplish effective and acceptable results.

However, the rapid growth of urban population worldwide (World Health Organization (WHO), 2013), spatial densification planning policies and lack of previous provision, not only may provoke a narrowing of park areas (Glickman, 1999; Polat and Yilmaz, 2013), but also can seriously threaten ecosystem services (Tyrväinen and Väänänen, 1998; Borgström et al., 2006). This unsustainable process should be controlled by nature conservation plans and green belt legislations. The consequences derived from the application of a range of possible planning approaches and regulatory policies lead us to envisage the evolution of multiple hypothetical future scenarios. The analysis of the plausible implications of each of them can give support to experts, planners and political decision-makers to understand the socio-economics and biophysical driving forces involved in this complex system (Alberti and Marzluff, 2004; Boyd, 2008) in order to elaborate guidelines or other kinds of legal mechanisms to mitigate the negative effects of urban development over biota and human beings.

Among other important functions, public open space planning allows local authorities to protect certain areas from the urbanisation process, and thus foster the formation of healthier urban environments. From this point of view, local and central governments play perhaps the most crucial regulatory role (more so than national governments or international organisations) in the control of land-use change in the longer term. These areas should be conceived as a valuable municipal resource, even though from a pure economic point of view, green areas can be perceived as a less tangible short-term benefits and savings good (Chen and Jim, 2008). In a cost-benefit analysis, their advantages are often underestimated (Kremen and Ostfeld, 2005) and complicated to quantify because of our incomplete knowledge (Vatn and Bromley, 1994). However, apart from the human dependence on nature (Bolund and Hunhammar, 1999), a utilitarian approach can be used here to justify the existence of these areas based on the favourable effects derived from the services offered and the fulfilment of immaterial necessities of the population (Maruani and Amit-Cohen, 2007).

However their distribution and location should be carefully studied by developing an adequate, long-term planning strategy. One possible option for maintaining a healthy urban environment is by reserving a collection of arbitrary areas to transform them into recreational parks. However, the time planning and geographic distribution of these spaces need careful consideration to ensure the quality and quantity of environmental services provided to the surrounding community (Forsyth and Mussachio, 2005). There is a lack of agreement about how to implement a given planning process and which measures should be selected. The most important points to discuss are:

- How to select adequate planning criteria.
- Deciding the most suitable size for the open space according to the current

and expected necessities.

- Where the open spaces should be located and how they should be accessed.
- The design of the potential activities for these areas according to different age groups and cultures.

A problem that arises when these issues are tackled is that there exist a variety of approaches with clear contradictory main goals. Among all of them, the present work follows a demand approach where the planning process should be based on attributes of the specific target population. The necessity of provision of a set of services defines the pressure over the available open space. This pressure can be measured by means of the analysis of features related to the urban population including size, density and distribution and also by collecting data related to subjective personal preferences for multiple non-homogeneous population groups. An advantage of this method is the simplicity, since it does not need further information about the ecological value of the selected areas.

To help in this complex decision-making process, an urban framework was created that could represent a paradigm of allocation of resources within a dynamical location-allocation decision-making simulator. The type of facilities aimed to spatially optimise is green areas, due to the level of importance these locations have for the society from multiple perspectives. The model represents a city in expansion, where a set of actions can be taken, modifying the dynamics and relationship of different elements that are included into the urban model. The model, and the subsequent strategies applied to optimise the location of this kind of facilities, is a very useful tool to further understand the dynamics and the interconnection of the processes implemented within it.

## 2.4 Modelling

Reality cannot be totally formalised. Modelling uses mechanisms such as theoretical abstractions and simplifications of real world entities to identify and replicate their essential features. These mechanisms allow us to find solutions to real-world problems where it is impossible or undesirable to experiment with the real system itself. For instance, very commonly the analysis of how a system responds, which is performed by designing studies and experiments, entails expensive costs that can be partially reduced by the use of this technique. Modelling is normally implemented in computer environments and can be applied from a twofold perspective: to predict the behaviour of a system or to enhance the understanding of its dynamical processes and emergent patterns before final decisions about implementation are made.

The process of modelling starts with an abstraction phase, where theory or observations based on the study of real historical information are taken and mapped into a formulation. These observations, that describe or test how the system evolves in time, can be analysed to find an optimised solution to the problem and/or to create feasible future projections. Finally, another mapping process is carried out where the solution previously found is translated into the real world in the form of a set of strategies, which may fulfil the expected future needs.

However, it is very challenging for the model to properly capture non-intuitive relationships among processes. One of the key factors, which contributes to the generation of successful modelling results, is the search for an appropriate level of aggregation. Here, 'level of aggregation' refers to the level of detail incorporated into the model, which is required for addressing the question of interest. Each additional degree of freedom represents a significant increment in the efforts required to structure and understand the model. If the level of detail is too low, the resulting model is not rich enough to synthesise an effective and realistic solution; if there is too much detail, the result is a cumbersome model, requiring perhaps immense levels of computational resources to run, and perhaps leading to only marginal benefits in outcomes. The granularity needed to transfer the proper amount of information from the real world into the model is an empirical question that depends on the underlying characteristics of the particular problem and the selected method to build, study and optimise the system (Friesz, 1985; Barker, 2014).

Normally, the 'micro-scale' in a model refers to discrete individual behaviours, possibly under the effects of several constraints. On the other hand the macro-scale is characterised by a limited number of aggregated variables, perhaps each representing a population of individuals. Since the representation of different scales alters the model predictions (Bovy and Jansen, 1983), selecting multiple levels and analysing their outcomes can be a desirable strategy. However, due to limitations in computational performance, not all aggregated levels can be considered.

### 2.4.1 Urban models

Cities can be defined as urban political units that play a significant role in socio-economic development on a global scale. However, they can also be the source of many problems, such as environmental pollution, traffic congestion, reduced open space, crime, diseases and social marginalisation (Lee, 2008). Their study as entities with their own dynamics is a very challenging task. Furthermore, urban and regional studies are, by nature, interdisciplinary areas of knowledge that can be faced from different angles. From this perspective, cities and regions can be analysed as non-linear spatially complex systems composed of many inter-related processes and socio-economic interactions and therefore can be difficult to understand in isolation (Jacobs, 1961). They exhibit characteristics such as fractal dimensionality, self-similarity across scale, self-organisation and emergence (Batty and Longley, 1994). Fractal dimensionality refers to the fact that urban patterns are arranged in regular distributions that are characterised by a fractal geometry nature, which has been reported and confirmed in multiple studies (Batty and Longley, 1994; Frankhauser et al., 2012; Frankhauser, 2015). Self-similarity is a property presented in systems with a degree of regularity in the generated patters. These regularities appear due to their scale-independent nature (Wolfram, 1994). The study of the self-organisation processes, which governs its evolution, involves multiple and highly complex levels of organisation. Changes occur over many years at an aggregate level of individual and collective choices, which mainly depend on the predominant characteristics of the time and on changes in external circumstances. This is the reason why real urban phenomena are so difficult to understand and replicate; the research community recently considered cities as one of the most complex elements ever created by human beings (Barredo et al., 2003) and even a science on its own (Batty, 2013).

In the context of urban development, the understanding of urban systems is a hard task, which is often full of uncertainty. With the support of modelling and simulation it is possible to reduce the level of uncertainty and facilitate a good degree of understanding. Modelling in this context can be viewed as a systematic approach to explore cities in different contexts. As such it can shed light on the major drivers of their evolution which could give rise to different plausible future alternatives. It can also help explain the topological structure of cities and their related growth dynamics.

If the model is focused on prediction and forecasting, a scenario-based approach can be used to provide multiple descriptions of alternative futures. Based on these scenarios, models can be also used as a tool to provide the necessary understanding to create feasible urban policies and explore the consequences of different strategies, addressing at the same time the problem of the inherent uncertainty of dealing with future projections (Wilkinson and Eidinow, 2008).

#### 2.4.1.1 Theoretical Models

Urban models can be designed to analyse real-world urban development processes in a concrete region with the use of real data and methods such as photogrammetry and laser sensing data (Santé et al., 2010; González et al., 2015), or can be designed as theoretical laboratories to deepen the understanding of urban concepts by the representation of a mix of elements of real cities and idealisations (Johnson-Laird, 1980). In general, this latter approach is focused on the creation of conceptual experiments within a simulated study area where a city is constrained within finite dimensions. Most such models are configured as a regular lattice, to simulate land use changes from non-urban to urban states, which can be used to test the viability of different urban development hypotheses or to explore a set of future scenarios.

The use of a theoretical model as a 'what if' laboratory has been explored by several researchers. Wu (2000) replicates some urban development configurations without including people's decision-making choices to distinguish between spontaneous or self-organising urban growth processes. Semboloni (2000a) analyses the suitability of a three dimensional Cellular Automata to model the dynamics of a monocentric city. Couclelis (1997), on the other hand, uses a theoretical model to pursue a more formal conceptualisation of regional models via *geo-algebra* mathematical expressions and the study of the proximal space concept in the context of linking Geographical Information Systems (GIS) data with these regional models. Meanwhile O'Sullivan (2002) uses a theoretical approach to explore the gentrification process in urban areas. The concept of gentrification can be defined as the renewal of certain urban areas produced by the settling of upper- or middle-income families or individuals in poor deteriorated urban areas. Liu and Phinn (2003) develop a Cellular Automata model of urban development incorporating fuzzy-set and fuzzy logic techniques in order to study the differences between an unconstrained urban area and a topological constraint scenario using a transportation network. Recently Sfa et al. (2015) uses this technique to construct a learning model to study land-use changes with the use of multiple advanced economic indicators.

### 2.4.2 Types of Urban models

A huge amount of varied models have been defined and developed to abstract urban phenomena which can be broadly classified into two different groups according to the process of abstraction used; a centralised top-down perspective and a decentralised bottom-up approach.

#### 2.4.2.1 Top-Down approach

In the 1950s-1960s the traditional planning methodology evolved into a more scientific perspective by the use of Large Scale Urban Models (LSUMs) (Lee, 1973). These models attempt to simulate the transformation of patterns of land use into an entire specific urban area. They can be characterised by their static and descriptive nature and, in general, by an unnecessarily level of complexity which makes them very hard to understand (Lee, 1994). The lack of transparency also contributes to difficulty in replicating results (Batty, 1979; Torrens, 2000). In summary, these models have failed to produce coherent results despite their great costs.

Top-down approaches are the result of conceiving cities as one type of social and natural phenomenon explainable by the use of traditional methods based on Urban Economic theory (Mills and MacKinnon, 1973), Urban Ecology (Dendrinos and Mullally, 1985) and Mathematical Geography (Wilson, 1972; Wilson and Kirkby, 1975; Allen and Sanglier, 1979).

Urban ecological modelling, for instance, proposes that the evolution of urban population dynamics towards an equilibrium state can explain the development of a city. This population can be considered, from a mathematical point of view, as a continuous function of time that shows growth and decline behaviour corresponding to the birth and death rates of individuals. By means of a deterministic strategy and simple associations at an aggregate level, the technique searches for global patterns at various time-scales. This approach mainly assumes a closed world, where no interaction with other cities occurs, and where variation in the population depends only on the current population (with no immigration or emigration, for example).

Another traditional and well-studied approach is based on urban economic models that use neoclassical microeconomic theory at an urban scale. The urban model proposed by Von Thünen (1966), derived from his agricultural land rent theory, the concentric model of Park et al. (1925) based on the spatial arrangement of Chicago city, and the urban spatial structure theory of Alonso (1964), all suggest that the major factors influencing the topological arrangement of residential areas are transport costs and distance to the city centre (Getis and Getis, 1966; Ullman, 1941). The approach assumes monocentricity and the maximization of a utility function (Molotch, 1976) which is considered the basis for the modern urbanization process and its associated economic growth. This approach also uses microeconomic concepts like perfect competition and total rationality, but it is limited by the simplification that spatial processes only result from individual choices and not from relationships between the elements involved.

#### 2.4.2.2 Bottom-up approach

Due to their difficulty with representing non-linear dynamical factors and bifurcation behaviours, top-down approaches cannot properly replicate urban phenomena (Lee, 1973; Su, 1998; Cheng, 2003). With a deeper and more complete understanding of urban phenomena, urban modelling evolved from a centralised approach characterised by top-down models, to a decentralised, dynamical and disaggregated perspective as an attempt to study cities as a highly complex system.

This new 'bottom-up' approach has developed alongside advances in computer processing speed and capacity, the emergence of complex systems theory, and the birth of more powerful artificial intelligence techniques like Matter-Element Models (MEMs) (Gong et al., 2012), Artificial Neural Networks (ANNs) (Park et al., 2011), Cellular Automata (CA) theory (Itami, 1994; Ligtenberg et al., 2001; Mahiny and Gholamalifard, 2007) and Genetic Algorithmss (GAs) (Porta et al., 2013) among others. These systems are based on the assumption that microlevel decisions in individual land parcels generate changes at macro-level scale, producing consequent spatial land use patterns that cannot otherwise be modelled or understood. This decentralised perspective also allows the application of new geometrical elements linked to chaotic behaviours and fractal structure (Wong and Fotheringham, 1990; Frankhauser et al., 2004).

There are growing examples of hybridisations of, and/or comparison between these two modelling paradigms. Wu and Webster (2000) proposed a combination of neoclassical urban economic theory with complex systems using a cellular GIS model for modelling urban processes, where transition rules are inferred from processes of the property market. Additionally, Schieritz and Milling (2003) conducted a study which compares a top-down approach, System Dynamics, with the bottom-up strategy Agent-Based Modelling, to model the dynamics of forests.

### 2.4.3 Alonso's Model

The theory of Alonso (1964) called *Urban Land Market Theory* is a static, uniform and continuous description of land use that was based on a refinement of the ideas of the

Von Thünen model, developed in the 19$^{\text{th}}$ century. The residential location model of Alonso is centred around the concept of *bid-rent function*, see Fig. 2.1, of a household, which can be defined as the maximum rent paid for a unit of land at a concrete distance from a unique CBD, which maintains a certain level of utility (Hoover and Giarratani, 1984). The concept of *utility* is based on the idea that satisfaction at individual level is associated with the consumption of certain goods or services.

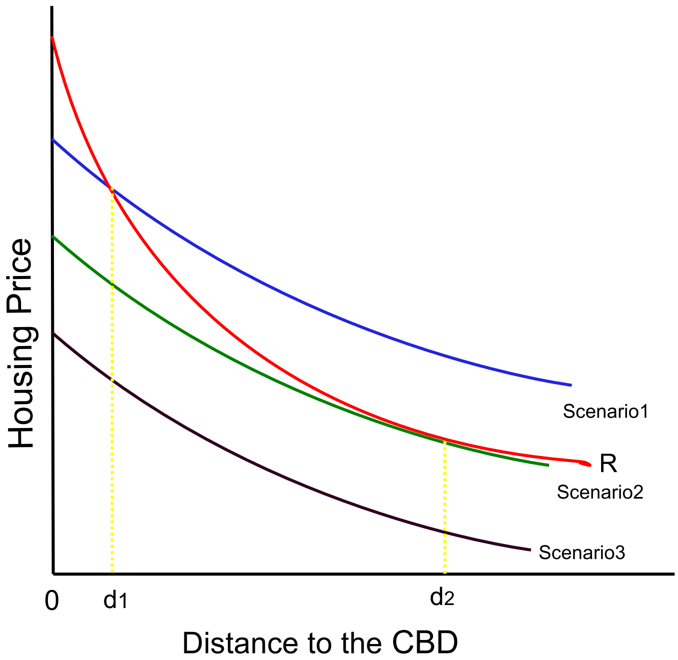



Figure 2.1: Bid-rent curves ($Scenario_1, Scenario_2, Scenario_3$) for one household, which depict a set of combinations of land prices and distances from the CBD ($d_1, d_2$) associated with a different level of utility. Among these options the household can select its residential location. The gradient of current rents in the city is given by $R$.

The micro-economic decision problem assumes that urban pattern formation is the consequence of individual urban residence choices in the selection of a new dwelling subject to budget constraints. Residential selection can be characterised by the generation of an economic competitive equilibrium for space between housing and commuting costs, which depends on the distance to the CBD. Daily residents commute to the CBD at a cost per unit distance. At this point the system achieves an equilibrium state, which yields the same utility level for all individuals. The approach assumes the existence of a densely radial road system, unemployment equal to zero, and the assumptions that all inhabitants earn the same income and have perfect knowledge of the market.

Formally, the choice of a household regarding the level of housing consumption and the location is the solution to the following utility function $U$, which is derived from consuming different combinations of goods, in this case $z$, $q$ and $t$, in such a way that:

$$Max : U = U(z, q, t) \tag{2.1}$$

where $z$ represents the amount of all other commodities, $q$ refers to land housing prices and $t$ is the distance from residence to work site. This maximisation is subject to the budget constraint.

In this scenario the population density and rent both tend to reduce with movement from the city centre to the suburbs, which can be approximated by a smoothly declining negative convex function. However, in real cases this population distribution tendency may fail if commute costs are high compared to wages, and also in the case that the distance from the household to the CBD is large. This effect can also cause segregation dynamics in the spatial distribution of low and high income groups (Brueckner et al., 1999).

The main limitations of this approach are that it is based on information that is incomplete and on unrealistic assumptions, such as those listed in the following points:

- In a real-world scenario, the resident market is always imperfect and affected by different forms of externalities (Batty, 2009; Cooke, 1983). Residents have to decide among a set of distinctive buildings, each of them with a varied range of heterogeneous properties.

- The concept of monetary maximization could be confounded by other factors of the landscape, which are likely to influence spatial patterns, such as accessibility of transport routes or heterogeneity of the landscape.

- It is difficult to collect real measurements of the utility function, and its real effects on the system, when important factors like congestion and commuting time are ignored. Political and governmental forces are also not included.

Because of these flaws the use of this utility function has been questioned (Hamilton and Röell, 1982).

- Most often, employment centres tend to move from the CBD to the suburbs due to the decentralisation of economic activities. Under these circumstances Alonso's predictions are significantly affected because residential decisions are partially subject to the expected workspace choice. Waddell et al. (2015) developed a residence-workplace location model specification to study the interdependence between residence and workplace choices for single-worker households and concludes that 80% of household decisions are conditioned to the job location.

### 2.4.4 Centricity

Spatial distribution of activities and facilities are important factors in urban morphology (Romanos, 1976). Centricity characteristics in urban areas are linked with a certain level of population and activity agglomeration, normally captured by the use of density functions. From the multiple types of possible dispositions, monocentricity is the arrangement that has been used the most in modelling urban patterns. This kind of layout is characterised by the existence of a unique point-wise CBD, physically located in the centre of the city where jobs and commercial activities are mainly concentrated and it is radially accessible from any other point of the city.

From a morphological formation perspective, monocentricity is considered a normal topological characteristic of industrial cities. Meanwhile multicentric arrangements, where each centre competes among the others, typically arise by a twofold mechanism. Firstly, cities mostly developed in a post-industrial age normally show spatial layouts with multiple CBDs that can be dispersedly developed in parallel by natural pattern formation and secondly it is also possible that by a merging process, several industrial cities close enough to each other may evolve into the same final spatial pattern.

In land-use change, the forces that cause centralisation and decentralisation phenomena, and, hence, changes in urban morphology, are called centrifugal and centripetal. Centripetal tendency is linked with the idea that residents concentrate in the centre of the city to get benefits from the service accessibility. On the other hand, centrifugal forces exert pressure on business and services to leave the centre of the city and settle in the suburban fringe (Pitzl, 2004). In real circumstances these forces do not affect the dynamics of the city independently; instead, multiple different centripetal and centrifugal forces are normally acting in an urban region simultaneously.

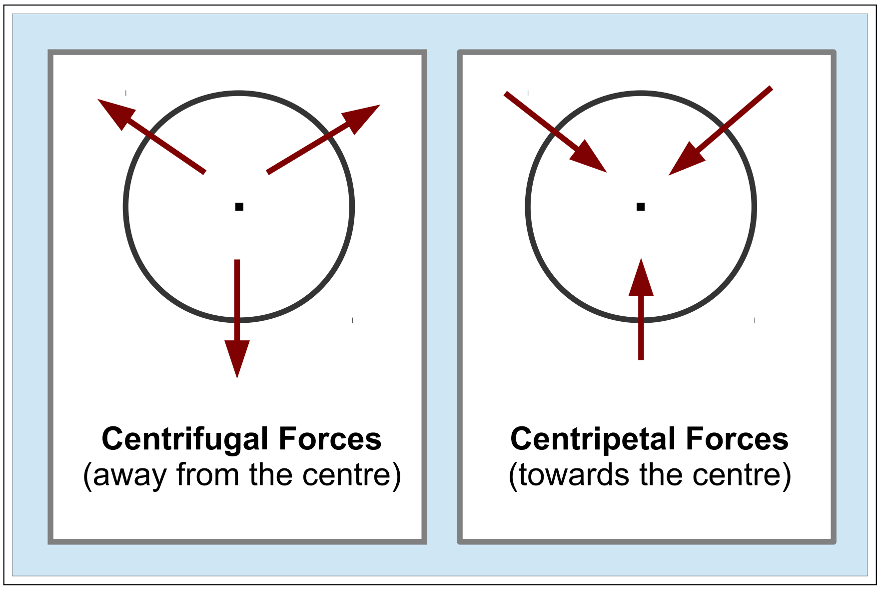



Figure 2.2: Different types of urban forces: centrifugal dynamics which cause movements to the outskirts of the city and centripetal dynamics that provoke the opposite behaviour, the concentration of activities towards the centre.

In modern times and in the majority of developed countries, most monocentric cities have experienced a suburbanisation and decentralisation process, which has caused the development of new metropolitan forms. This effect has derived in a general tendency towards the evolution from monocentric structures to polycentric dispersed cores, where centripetal forces are replaced by centrifugal ones. These dynamics lead to a population density decay in the most crowded inner areas of the city. Meanwhile, in outer suburbs, the density tends to rise as the city grows (Berry, 1977, p. 108). This pattern originates from the desire of people to reduce their commuting times, and from the displacement of job opportunities to other areas of the city. The latter comes from flexibility in the production, where businesses benefits from lower land costs and location, avoiding competition among CBDs in a single city. The result is a multicentre city with multiple cores and various housing price gradients.

Since spatial centricity was an element included in the first urban models (Mills, 1967; Muth, 1969) the monocentric assumption has received noticeable empirical support (Clark, 1951; Frankhauser, 1998) and still partially dominates current urban models (McMillen, 2006). However, numerous authors state that this feature could be a limitation in capturing the fundamental characteristics of the city (Wheaton, 1979; Berry and Kim, 1993) because of its simplicity and the existence of various market failures including the interlocation relationships between agents (Irwin and Bockstael, 2002), among other factors. All of these flaws make this framework insufficient to explain real world complex spatial structures like mixed space and scattered development (Caruso et al., 2007).

The application of monocentric assumptions to urban models characterised by multiple economic cores produces a distorted panorama of the distribution of urban elements (Romanos, 1977). However it is a useful analytical tool and the basis to be applied in the modelling of polycentric cities (McMillen, 2006) because it shares the same principle, namely that agents' choices determines the urban spatial distribution, and it is sufficiently robust to explain other kind of spatial patterns, including single and multiple clusters and dispersion (Irwin and Bockstael, 2002). Even if nowadays it is assumed that the island-city no longer exists (Fontaine, 2010), this simplification is still used because it avoids the need to include interurban factors into the model. Apart from that, Bertaud (2004) states that, in the real world, urban areas are not purely monocentric or polycentric.

### 2.4.5 Limitations of Modelling as a technique

In this section a list of drawbacks that the model has to face due to different characteristics of the system is illustrated.

**Time horizon of the model.** Modelling spatial processes and dynamics require us to envision possible future conditions. In this regard, the defined time horizon of the model is a crucial parameter to take into consideration. The reason behind this is that it is not possible to construct reliable predictive long-term models simply by extrapolating from current characteristics of the system, since this is inherently biased, boosting the likelihood of only one future scenario based on inaccurate assumptions over plausible future evolution.

**The predictability problem.** The predictability problem states that a massive number of possible futures exist (Greeuw et al., 2000). A process can evolve in different ways due to associated stochastic elements that cannot be controlled and foreseen. However, in the real world only one present and one future emerge. Validating models against a single reality may not be accurate enough to derive acceptable assessments of their rationality (Brown et al., 2005). These non-linearities and variability can be tackled by deciding which of those patterns is the most representative for the goals of the agents (Batty and Torrens, 2005).

**The inference problem.** Fotheringham and Brunsdon (2004) stated that in spatial data analysis and spatial modelling, a given pattern can be the final result of different processes, and that different types of evidence result in multiple kinds of inference that can be also valid.

**Level of detail.** The Bonini's paradox (Bonini, 1963) consists of the accuracy - comprehensibility trade-off that said that, as a simulation model moves toward representing the full complexity of a real system, it forgets its comprehensibility and transparency. To achieve a high level of sophistication, the required assumptions become so complex and their inter-relationships so obscure, that the model may not be easier to understand than the real system itself.

**Calibration of the model.** The calibration of a model is normally carried out using historical data, however land use models often include randomness to simulate complex processes. Bifurcation and emergence are intrinsic characteristics of urban dynamics (Batty, 2007) and results are generally path dependent (different outcomes can be generated by the same model (Brown et al., 2005)), which implies that a given outcome may represent plausible dynamics even if it does not match the actual land use change. Consequently some authors argue that accuracy assessment is not the most appropriate strategy to measure the quality of the simulation results (Parker and Meretsky, 2004; Power et al., 2001; Remmel and Csillag, 2003). An alternative method to assess the suitability of a model is pattern based measurements including patch characteristics (Riitters et al., 1995), polygon matching (Power et al., 2001) or fractal analysis (Frankhauser, 2004).

## 2.5 Cellular Automata

CA were proposed in the late 1940s by John von Neumann and Stanislaw Ulam with the name of 'cellular spaces' for discrete space-time representation of complex dynamical problems which obey their local physics (Von Neumann, 1966). In this work, the ambitious aim of the authors was to create a system that was capable of simulating idealised biological systems, with a special focus on modelling self-reproduction phenomena. For that purpose, they use a set of 29 different states and a neighbourhood composed by four adjacent cells.

In the context of complex analysis, the simple ecological model developed by the Cambridge mathematician Conway (1970) called the 'Game of Life' is the perfect example of the use of a CA for representing artificial life. In this work, he defined a two-dimensional CA with two states (dead or alive) with the main goal of finding a simple rule that generates complex behaviour. With the definition of only two transition rules, the model turns out to have an unexpected potential for representing rich behaviour. Another remarkable study of the abilities of a single one-dimensional CA was the one created by Wolfram (1984), who also demonstrated that CA has the capability of modelling complex natural phenomena. By studying the rules of an one-dimensional cellular automaton in detail, it was concluded that this system can show stability, stochastic instability or chaotic behaviour.

CA can be applied to a huge range of areas including natural sciences, mathematics, and computer science. In this regard, Lafe (2012) groups the areas of applicability of CA into four main categories:

- as powerful computational engines (Conway, 1970).

- emulation of discrete dynamical systems in simulation.
- as conceptual tools to analyse pattern formation and complexity (Wolfram, 1984).
- as the basis for modelling fundamental physic behaviours, such as ising models (model of ferromagnetism in statistical mechanics) (Creutz, 1986) or turbulence phenomena (Chen et al., 1988).

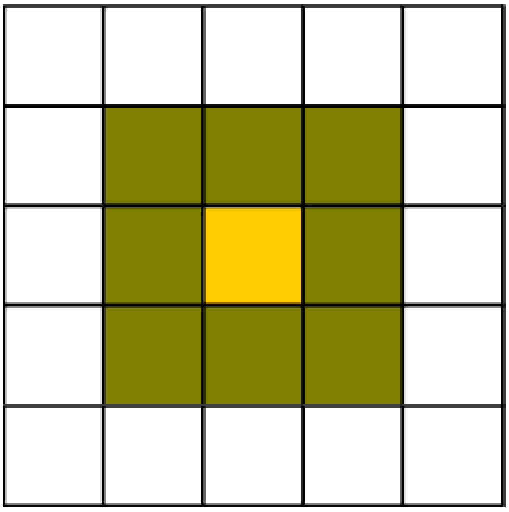
In general terms, a CA consists of a regular discrete lattice of cells, where each of them has associated an internal state selected from a finite set of possible values. The dynamics are managed by a set of decision rules, which defines how cells can evolve from one state to another. These dynamics are based on the assumption that, by means of local aggregated interactions generated by these simple transition rules, the model is capable of representing complex phenomena. This aspect that characterises complex systems, in which a small number of local actions are capable of generating complexity by an aggregate process, is called emergence.

The non-linearity nature of the iterative evolutionary process of a CA can lead to the generation of regular fractal patterns (Batty and Longley, 1994; Longley and Mesev, 2000), which is a frequent characteristic in fields like urban environments. By the use of this technique, experts can therefore gain insight into the appearance of such complex spatial phenomena in the real system. Meanwhile in this context, the CA is an excellent tool to explore complexity of this kind at multiple time-scales.

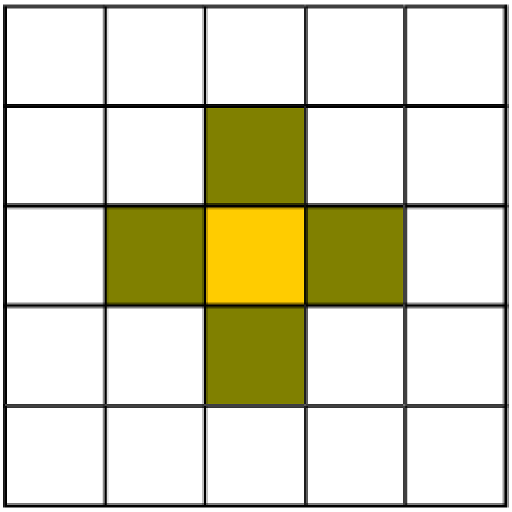
A traditional CA can be defined as a finite-state machine, in which the state of each cell at time $t+1$ is dependent on two factors: its state at time $t$ and its neighbours according to a set of transition rules. Formally, this can be expressed as follows:

$$S^{t+1} = f(S^t, \Omega, T) \tag{2.2}$$

where $S^t$ and $S^{t+1}$ are the sets of all possible states at time $t$ and $t+1$ respectively, and $\Omega$ is the neighbourhood of all cells. Both aspects provide input values for the transition function $T$ that defines the change of the state from $t$ to $t+1$.



(a) Visual representation of the Moore's neighbourhood.



(b) Illustration of the cells arranged according to the Von Neumann's neighbourhood.

A neighbourhood can be defined as the cell itself plus a certain configuration of cells around the examined cell at some distance (Torrens, 2000). The size of the neighbourhood determines the amount of surface information that is considered in the neighbourhood process. In a two-dimensional CA, the two more famous neighbourhood are the Moore's and the Von Neumann's neighbourhood represented in figs. 2.3a and 2.3b.

Transition rules are the key element in the evolution of a CA. They are mathematical expressions that define the changes in behaviour of each cell in the lattice. Usually represented by a set of 'IF–THEN' statements, transition rules can be static, probabilistic or use more advanced methods to evolve. Advanced versions of non-linear transition rules are capable of allocating self-modifying cells where they automatically evolve towards more efficient states using optimisation techniques like GA (Mitchell, 1998), Ant Colony Optimisation (ACO) (Liu et al., 2007), Support Vector Machine (SVM) (Yang et al., 2004) or Markov Chain Analysis (MCA) (Mitsova et al., 2011). The use of EA to calibrate them has been also investigated in Shan et al. (2008).

The CA process starts with a predefined initial configuration of cell states. This configuration refers to the spatial patterns over the entire lattice defined by these local states. Traditionally, in discrete time-step systems, when the state of every cell in the lattice is updated, these changes are synchronous. This means uniformity in the update process of the cells.

The advantages of this technique are numerous. In the next list, some of the most important factors are highlighted:

- Transition rules are intuitive and easy to understand.
- It can be easily implemented on computer platforms.
- Its configuration and evolution are independent of any given statistical distribution.
- The technique is good at performing spatial simulations; the process and its outcomes can be visualised immediately with spatial analysis tools like GIS, raster-based data and remote sensing (Couclelis, 1997; Wagner, 1997). These factors allow the inclusion of high-resolution spatial data.
- Due to its computational structure, CA well supports efficient parallelisation (Gibson et al., 2015), facilitating their use in large-scale simulations of scientific models or image processing.

Meanwhile, Ward et al. (2000) have questioned the validity of the global patterns generated by a CA. They highlighted the necessity to extend the CA modelling mechanisms based on self-organisation processes at a local level to introduce some broad-scale factors to constrain and modify these local aggregated dynamics.

### 2.5.1 Cellular Automata in Urban Scenarios

CA have been shown to be powerful qualitative tools for modelling, not only urban phenomena (Batty et al., 1997; Benenson and Torrens, 2004; Ligmann-Zielinska et al., 2005; Matthews et al., 2007; Santé et al., 2010) but also for addressing other spatial simulation problems such as robot path planning (Ferreira et al., 2014), pedestrian modelling (Crociani et al., 2015) or fluid dynamics (Lafe, 2012). Tobler (1979) was the first who used CA techniques to model geographical phenomena. In the concrete field of urban development, CA was also used to explore the driving forces of urban development (Arsanjani et al., 2013; Leao et al., 2004), analyse urban growth patterns (Aguilera et al., 2011; Moghadam and Helbich, 2013), support policy analysis and land use planning (Geertman and Stillwell, 2004) and explore scenarios for future development (Barredo et al., 2003; Engelen et al., 2003). In these two-dimensional models, changes in land-use are depicted mostly as urban cells which spread out from a determined point to adjacent neighbouring points (Santé et al., 2010).

To make CA applicable for land use modelling, the strictly defined elements are frequently relaxed. In this regard, urban growth simulations within a CA context are normally limited by the size of the predefined area under consideration. To avoid this constraint and create an unlimited area of study, an infinite line can be represented by a circle if the framework is configured in a single dimension or, in the case of two-dimensional scenario, an infinite plane can be represented by a torus, a finite rectangle whose opposite edges are connected. Other methods use a variable grid to aggregate more remote areas by means of generating field approximations. When required, the model is able to enlarge a specific neighbourhood to include cells at all distances by using a hierarchical representation of space (White, 2006).

Another element to consider is the dimensions of the defined grid. In urban development, CA is most commonly defined as a two-dimensional lattice, however Semboloni (2000a) analysed the behaviour of a city using a three-dimensional CA as an extension that permits the inclusion of richer parameters in the model that are problematic to consider in a normal CA. As such, the model represents population density variations and mix of land uses by including information that represents land height, in the case of non developed land or number of floors for urban areas.

The characteristics of cells have also been investigated since raster-based data shows some limitations in term of geographical representation, even when they are fitted with high-resolution data (Benenson and Torrens, 2004). A raster can be defined by a matrix of cells or pixels arranged into a structure of rows and columns forming a grid. The raster representation of an area eases the definition of the neighbourhood and its subsequent relationships. Irregular tessellations are more appropriate for real representation of the land.

Traditionally the shape of cells within the grid is square, however Iovine et al. (2005) proposed the use of hexagonal cells in order to obtain a more homogeneous neighbourhood layout. The use of non-identical cells was covered by O'Sullivan (2002) who proposed a *graph-based* CA, where each node with irregular spatial tessellation was an entity of interest located in a defined point of the lattice. In Shi and Pang (2000) this option was investigated along with the use of *Voronoi polygons*. A Voronoi polygon is a polygon whose extension is made up by all points in the plane which are closer to a given particular point in the lattice than to any other. The use of this type of geometrical structure was also included in other urban studies (Semboloni, 2000b; Shiyuan and Deren, 2004). Another example of the use of advanced geometrical techniques is the case of the proposed *Delaunay triangulations* (O'Sullivan, 2001a,b). Finally, the definition of *cadastral parcels* instead of regular cells, which leads to a more realistic representation of urban phenomena was analysed by Stevens and Dragićević (2007).

The cell space has been also extended to include non-uniform features with land attributes such as slope, existing land cover, elevation, accessibility and the angle of the area. Consequently, in this case the cell space is not uniform, which entails that some cells are more suitable for certain land uses than others (Wahyudi and Liu, 2013).

The selected neighbourhood defines the amount of land that can exert influence on the parcel under consideration, and its effect depends normally on the type of land-use and the distances between these parcels of land. In this context, the previous formulations significantly affect the definition of each neighbourhood and their interactions, which may require significant revision. For instance, the concept of neighbourhood could also be defined as a non-static parameter when it is used in a graph representation (O'Sullivan, 2009; Barreira-González et al., 2015). In this case the neighbourhood can be defined by the adjacent units within a specified distance or by a Voronoi spatial model.

In a traditional CA, only directly adjacent cells are included in the neighbourhood. In the real world, different land-uses are affected by dynamics that may happen at greater distances, even if typically this effect decreases with distance. Hence, larger neighbourhood configurations have been used to model these kinds of land-use relationships (White and Engelen, 1993; Barredo et al., 2004; van Vliet et al., 2009). However, such configurations include more information, so the required computation time to deal with this extra amount of information may increase dramatically, because the number of cell-to-cell relations grows with distance. To incorporate the effects derived from operating over larger distances, Gravity-Based Regional models (White and Engelen, 2000) or a more complete hierarchical conceptualisation of the space (Andersson et al., 2002) can be used.

States can be also differently defined, from the most simple approach as binary values which can represent urban and non-urban areas, to qualitative values of multiple land use types or quantitative values that could represent population density (Li et al., 2003), degree of development (Yeh and Li, 2002), the value of buildings (Cecchini and Rizzi, 2001), or a vector that includes several attributes (Portugali and Benenson, 1995).

Regarding the nature of the transition rules, the approaches most commonly used are non-linear or probabilistic. The application of these kinds of strategies assume that new developments are transformed from available land according to a certain probability. On the other hand, Wu (1996) investigates a linguistic simulation approach, where a set of fuzzy transition rules are aimed at selecting the most appropriate type of land-use using the highest grade of membership of the fuzzy set. Li and Yeh (2002) proposes the use of a three-layer ANN to calculate the conversion probabilities that allow the evolution of multiple land uses types by using observation data collected from remote sensing sources, and White et al. (1997) use the transition potentials from previously calculated calibration data. These implementations reflect more realistically that different land developers in different locations can vary the way they apply their financial resources to the construction of urban areas, and how they react to the different market conditions. Such an extension breaks the spatial symmetry, generating the development of clusters with irregular edges that are much more close to the characteristics observed in real cities. Apart from the multiple range of possible definitions, Brown et al. (2005) highlight the problem that transition rules can be differently adjusted according to alternative explanations of the same phenomena.

The spatial heterogeneity represented in different parts of the city should be addressed by different transition rules. Besides this, spatial heterogeneity has some effects on the morphology of the neighbourhoods. Realistic modelling should be described with neighbourhoods of different shapes and sizes (White and Engelen, 1993). These extensions allow the model to better capture the spatial interactions among urban structures.

Finally, variations in the duration of the time steps can be also studied. Irregular time steps, varying with the cell under consideration, has been analysed in Stevens and Dragićević (2007). Another less frequent variant is the use of variable time steps to model events of different lengths of time (Couclelis, 1997). Couclelis (1984) critiqued the inability of CA to deal with stochastic behaviour due to the updating synchronicity of the urban environment at each iterative step. Real cities violate this assumption due to their inner chaotic behaviour.

### 2.5.2 Examples of Cellular Automata-based Urban Models

The SLEUTH model, proposed by Clarke et al. (1994), is an adaptive CA-based urban growth prediction model that can be considered the most widely used by the urban community, due to the fact that it has been applied to over 66 different cities and regions(Clarke, 2008). This broad level of acceptance is mainly due to the incorporation of advanced characteristics like back-casting, the use of historical data for calibration of variables, a successful implementation at regional scale and its capability of simulating both urban and non-urban dynamics.

The framework comprises two modules, the Urban Growth Model (UGM) and the Landcover Deltatron Model (LCD) that can simulate land-use change dynamics. The model also includes thirteen metrics used to assess the best fit growth coefficients. SLEUTH allows the definition of multiple features such as slope, land-use types, urban extent, exclusion, transportation accessibility and hill-shade. The first study involving the SLEUTH model was focused on the historical urban growth development of San Francisco Bay and it was implemented by Clarke and Hoppen (1997). Silva and Clarke (2005); Wu et al. (2009); Akın et al. (2014) are some examples of modern applications of this model.

An advanced version of the SLEUTH model, called SLEUTH-3r was introduced by Jantz et al. (2010). The approach improves the performance in terms of memory consumption and its applicability by enhancing the calibration process with the inclusion of new fit statistics. Jantz identified a bias behaviour towards edge growth which may affect the dispersed development in detailed resolution data. This advanced version also implements a functionality which allows the model to identify the most promising areas where urban growth may occur. This advanced model has been demonstrated to be five times more computationally efficient than its previous version, reducing memory usage by 65%. Examples of the application of SLEUTH-3r include Belyea and Terando (2010) for the South Atlantic Migratory Bird Initiative (SAMBI) region or Jawarneh et al. (2015) for the study of future land development in central Arkansas (USA).

Metronamica (van Delden et al., 2005) is another dynamical CA-based land use-transportation model developed in the Netherlands by the Regional Integration Knowledge System (RIKS) company. Its main objective is to explore the effects of multiple socio-economic and physical policy configurations to aid in a broad range of decision-making processes. The model is composed by three main elements that cover distance decay functions, GIS integration and constrained cell transitions that are defined by calculating a ranked score for each cell. This framework has been proven to be very generic and flexible for a variety of spatial and temporal scales levels. Metronamica was applied to different regions (Daneke, 2013; Linke, 2008) and compared with the SLEUTH model in (Kim and Batty, 2011).

Another recent CA-based urban growth model that has appeared in the literature is iCity. iCity, that stands for Irregular City, was developed by Stevens et al. (2007). It is a predictive urban growth model that extends the functionality from a traditional CA by including an irregular lattice and an asynchronous updating behaviour of the cells. iCity was also the basis of another urban study, where a similar implementation supported the capability of the model to assist urban planners (Jjumba and Dragićević, 2012).

SimLand is a CA model used for simulation purposes based on the integration of a Multi-criteria Evaluation (MCE) model to derive behaviour-oriented rules for the transition among states that includes Analytical Hierarchy Processes (AHPs) (Saaty, 2008). The model was introduced by Wu (1998) and the most noticeable strengths include easier retrieval of spatial data, the use of GIS technologies to map multiple data layers and the integration of multicriteria evaluation methods. Based on that, the model achieves a more realistic definition of its AHP-derived transition rules.

The Canadian Agent-Based model system called Integrated Land Use, Transportation, Environment (ILUTE) was developed by Miller et al. (2004) as a tool to replicate the evolution of an entire urban region over a defined time horizon. The model analyses the demand-supply interactions of three type of agents namely landowners, households and developers in a residential and commercial real estate market. The ILUTE model was successfully applied in different scenarios (Salvini and Miller, 2005; Chingcuanco and Miller, 2012). Another Agent-Based residential model designed to study the urban land market was the one developed by Parker and Filatova (2008) which was focused on the interactions and relationships created between multiple trading agents including household agents, developers and rural-land owners in the pursue of a profit maximization objective.

Other models are designed applying hybrid strategies that combine a CA approach with other advances techniques. In this regard, the work of Al-Kheder et al. (2008) applies a fuzzy inference process to provide common semantic and linguistic knowledge to the urban growth model, simplifing the set of transition rules. Using satellite images the framework guides a CA to determine the potentiality of each pixel to be urbanised. Wang and Li (2011) applied a Radial Basis Function Neural Network Model (RBFNN) to generate the conversion probabilities from the initial urban cell to each objective land use type. The approach faces a problem due to the limitation of the ANN technique, which cannot explicitly identify the contribution of each variable. The consequence is that less important variables may be included into the model. To overcome this limitation, Kocabas and Dragicevic (2007) also investigates a Bayesian Network variant of the model, where land use factors and probabilities can be clearly interpreted in comparison with the Neural Networks weights. The calibration of the weights used in the transition rules were also determined using an adaptive Monte Carlo sampling taking information from historical data of the urban growth (Chen et al., 2002).

### 2.5.3 Calibration and Validation of Cellular Automata models.

Calibration and validation of CA models have been two of the most critical issues in order to consider a CA as a reliable method for urban growth modelling. Calibration can be defined as the process of generating the best fit for the model in order to recreate a determined behaviour that mimics certain dynamics of the objective system. However, it is a crucial factor in such simulations in which the parameter values or weights are evolved, so that realistic results are included into the model (Wu, 2000).

CA logic is mostly based on the values of the transition rules. Most of the times these values are assigned relying on an intuitive understanding of the changing process of the cell status. The problem is that the number of possible transition rules that give rise to structured global patterns is almost unlimited. Identifying an appropriate rule among a huge amount of alternatives is very complicated. Some attempts have been investigated to formalise a reliable procedure to apply the CA in an urban scenario. As such, Takeyama and Couclelis (1997) proposed a mathematical geo-algebra language for CA that is capable of formalising and generalising multiple dynamic spatial models. Xie (1996) also developed the Dynamical Urban Evolutionary Modelling (DUEM) model as a generic paradigm to construct dynamic geographic models to contribute to the understanding of urban phenomena like the simulation of different forms or simulating sprawl phenomena with the final goal to aid in decision-making processes. The model was applied to multiple real systems like the city of Buffalo (Batty and Xie, 1994) and Lanzhou in China (Xu et al., 2007).

Historical empirical data can be used as a tool to find suitable parameters values using techniques like MCE (Wu and Webster, 1998). Calibration has been also addressed by intensive computation simulations, using iterative processes with allow the model to run and test with different combinations of parameter values (Clarke and Gaydos, 1998), applying Fuzzy Logic (FL) (Liu and Phinn, 2003), training an ANN (Li and Yeh, 2001) or using GAs (Li et al., 2013). The two latter methods are capable of automatically generating the parameter values, even if these values may be complicated to understand and interpret.

On the other hand, to create a validation process for CA models is still a challenge. The validation procedure should be able to measure whether the model can capture the general trends and dynamics of urban land-use or not. The method most often used is to carry out a visual comparison to confirm the validity of the simulation results. This performance should be analysed only in relation to the specific pursued goals of the specific problem under consideration. In this regard, Clarke and Gaydos (1998) used four statistical indices to assess the accuracy of the simulation, Li and Yeh (2001) explores a conversion matrix and Li et al. (2013) uses the coefficients resulted from different density functions.

An additional possible method is the use of the trends and patterns retrieved from other models to measure the validity of an external model. However, comparing the performance among planners is generally a very hard task. Furthermore, most of the factors that hinder comparison are out of the scope of the prediction process, such as focusing on different operators and criteria to characterise the same reality, facing missing, incomplete or inconsistent empirical data in the form of inaccurate diagrams and poorly organised descriptions, or dealing with multiple optimisation procedures and configurations, and hardware characteristics.

## 2.6 Agent-Based Modelling

Agent-Based Model (ABM), formulated in early 1970, is a powerful simulation modelling technique that is composed by a collection of heterogeneous entities called agents with decision-making behaviour that can be arranged in different levels of organisation (Ferber, 1999; Parker et al., 2003). These autonomous entities can represent an heterogeneous community, capable of interacting among them and with their own environment locally, acting and adapting their behaviour accordingly. This local characteristic refers to normally agents do not interact indiscriminately with the rest of the agents, but only with their neighbours.

Each agent, that can exist at different scales or levels of granularity, is capable of assessing its current situation and making decisions according to a set of rules that defines their behaviour. These rules, that can be static or dynamic, allow agents to perform a change of state, phase or activity that the agent is engaged in. Mubasher and Jaffry (2015) investigate a case of dynamical behaviour in agents' rules, where five different agent profiles evolve through time by using the feedback produced by the environment in a Cognitive Driver Model (CDM).

The ABM technique has been used to understand the interconnections, interdependences and feedbacks created among a set of individual entities in order to fulfil their goals within a designed environment (Rounsevell et al., 2012). By the study of these simple repetitive and aggregated interactions, it is possible to interpret future socio-economical and ecological trends and global patterns in the form of distributions and their correlations (Gimblett et al., 2001). The knowledge gathered can be translated by experts in order to create a set of feasible strategies. Finally these strategies would be applied to the real-world problem. ABM has been widely applied to simulate complex systems in different fields ranging from economics (Hokamp, 2014; Hamill and Gilbert, 2015) to language (Pitt and Mamdani, 1999), social science (Ronald et al., 2012), ecology (Bert et al., 2011), biology (Bogle and Dunbar, 2010) and engineering (Pipattanasomporn et al., 2009).

The major advantage of this technique is that it operates from a bottom-up perspective by providing a mean to incorporate individual-level dynamics, which confers a natural description of the system. This feature allows the exploration of dynamics out of the scope of pure mathematical techniques (Epstein and Axtell, 1996; Axelrod, 1997). Multiple agents are able to interact leads to a highly dynamic and non-deterministic environment able to capture emergent phenomena and self-organisation patterns (Heath et al., 2009). Furthermore, the model can be considered an across-level technique since changes at the individual agent level may cause unexpected behaviours in the entire system. Remarkable is also its capability of integrating a broad set of knowledge sources and the ability of supporting the representation of different rich individual profiles that can be parametrised from quantitative surveys. Benenson and Torrens (2006) also add that the use of ABS does not require the full understanding of the phenomena under consideration.

Agents can be assisted by heuristic methods in learning from previous choices in order to achieve more efficient decisions to accomplish their objectives in the future. Some techniques have demonstrated significant potential including evolutionary algorithms (Manson, 2005; Bennett and Tang, 2006) and reinforcement learning (Bone and Dragicevic, 2010). In this regard, Tang (2008) concludes that evolutionary approaches are specially suited for gathering knowledge from a group of agents.

Validation and parametrisation is one of the main research challenges in the field of ABM. Commonly, the resulting models can strongly depend on the initial conditions of the individual entities and their simulated environment, and the way processes and events are scheduled within different submodels. As a consequence, most often formal ABMs descriptions are descriptively lengthy and incomplete, which makes replicability very hard or even impossible (Grimm and Railsback, 2012) and avoids transfereability of knowledge between models (Hales et al., 2003). From this perspective, several works address the pursuit of designing a standardisation process in the design of ABM. In this regard, Boero and Squazzoni (2005) developed a methodology to be followed in order to perform an empirical validation of ABMs. They created a taxonomy of models and a subsequent classification according to their intrinsic nature. They divide the models into three categories namely case-base models, typifications and theoretical abstractions. Afterwards, they analyse the empirical data needed and the validation strategies available in each case, creating a 'best practices' list for each of them. Richiardi et al. (2006) proposed a three-stage protocol to define the methodological standards for agent-based simulations in the social and economic field. They use questionnaires and the interaction of working groups to be able to design an initial set of recommendations to lead to a methodological standard for agent-based simulations. Other approaches are less ambitious and are based on the idea that replicability by different models and on several platforms can be seen as a way of being more confident of the veracity of the results (Gilbert and Troitzsch, 2005).

### 2.6.1 Agent-Based Modelling in Urban Scenarios

In the field of urban development, agents could represent civic or local government decision-makers, real estate developers, farmers or a recent arrived immigrant making iteratively endogenous decisions about the sale, purchase and development of patches of land. These individuals can be also grouped together in organisations and interest groups where they may exert extra influence over the system (Tian et al., 2011).

Based on simple actions and interactions, such systems are broadly able to model urban growth, cooperation processes and a given scenario with a concrete land-type distribution developed over time. The inferred plausible properties and patterns can give support to decision-makers and policy-makers to understand these complex social process. Future socio-ecological trends like the emergence of urban patterns can be derived from their cooperational behaviour, their endogenous economical choices and their direct and indirect interactions. For some authors ABM is considered more like a mechanism to explore different types of dynamics of a system rather than generating predictive results (Brown et al., 2006).

In the past 40 years and since the first studies were published (Schelling, 1971; Sakoda, 1971), ABMs have been commonly used in the geographical and land-use sciences. More recently this technique was applied to compare micro-macro urban phenomena (Schieritz and Milling, 2003), to study different modelling spatial patterns to maximise society's well-being by capturing Thünen's assumptions (Sasaki and Box, 2003), to simulate residential relocation processes and price evolution in housing markets in a not spatially explicit model where agents' decisions are based on perceptions probabilities (Ettema, 2011), to mimic traffic simulations within cities (Balmer et al., 2004) and also in the field of air traffic flow management (Agogino and Tumer, 2012) in order to analyse the complex interaction developed by the aircraft, airports and traffic controllers or in other areas of ecology such as population dynamics and movements of social animals, dynamics of plant population and ecosystems (Grimm and Railsback, 2013).

Torrens and Benenson (2005) conclude that the use of ABMs in conjunction with CA produces better simulation results. Together, these techniques create a conceptual framework which incorporates human decision making and spatially explicit representation of land use in addition to all their combined driving factors. ABMs, along with CA taking the role of representing land-use change dynamics, have been applied broadly to study land-use and land-cover change (Parker et al., 2003) and urban growth phenomena (Batty, 2007; Matthews et al., 2007) like rapid urbanisation processes (Xie et al., 2007) in the analysis of crowd dynamics (Mordvintsev et al., 2014; Luo et al., 2010) and as a tool to aid stakeholders for decision-making purposes and create suitable spatial planning policies with the use of future plausible scenarios (Ligmann-Zielinska and Jankowski, 2007). Mentionable is its use in the analysis of residential selection within a non-stationary housing market (Devisch et al., 2009; Parker and Filatova, 2008; Otter et al., 2001; Brown and Robinson, 2006). Zellner (2007) studied the water-use and land-use patterns derived from different policy alternatives by means of the Water-Use and Land-Use Model (WULUM). Filatova et al. (2009) applied these tools to analyse how land prices affect the behaviour or households in an urban environment. Robinson and Brown (2009) proposed a GIS-agent-based model called the Dynamic Ecological Exurban Development (DEED) model that evaluates the effects of lot-size zoning caused by municipal land-acquisition policies on available forest stands by the use of hypothetical scenarios. Finally, Miller et al. (2004) studied the role of transportation in the evolution of an urban region and Ehlert and Rothkrantz (2001) the evolution of human-like driving behaviour in agents which can exhibit different driving styles.

### 2.6.2 Repast Simphony as a modelling tool

Repast Simphony (RS) (REcursive Porous Agent Simulation Toolkit) is an agent-based modelling and simulation toolkit commonly used in the CA-ABM community.

The framework is free, open access and it has been under continuous development for about 15 years. The version 2.0, used in this thesis, was released on March 5, 2012 as a second-generation environment that builds upon the previous Repast 3 library (North et al., 2006). The framework uses Eclipse as its primary development environment that runs under Microsoft Windows, Apple Mac OS X, and Linux.

With a strong focus on well-factored abstractions, RS is the result of a highly modular plug-in architecture that allows the interconnection of individual components including networks, logging, and time scheduling. Each module can be connected or disconnected as required. Its design is structured to separate specification, execution, data storage, and visualisation. 2D and 3D OpenGL-based allow that the same agent to be displayed simultaneously in a variety of topologies. The framework also includes the integration with third-party applications, such as the R statistical package (Hornik, 2012), JUNG network analysis system (O'Madadhain et al., 2005) and the free GIS software GRASS (Neteler and Mitasova, 2013). Data collection is designed to gather and store information in running time typically from both, the state of the simulation and for each agent at each time step.

Previous publications have validated the software in different contexts (Artel et al., 2011; Parry et al., 2006; Griffin and Stanish, 2007).

## 2.7 Evolutionary Algorithms

EAs are popular and powerful stochastic metaheuristics that mimic the behaviour of natural selection postulated by the English naturalist Charles Darwin in the 19$^{th}$ Century (Darwin, 1861). This strategy, not based on neighbour search, is established under the assumption that nature evolves by the course of new generations, preserving the specimens more suited to their environment while the unfavourable individuals tend to perish. To accomplish that behaviour, the method uses a set of individuals, each encoding a possible solutions of the problem. This structure of the encoding corresponds to the biological genotype and each of the possible realisations to the phenotype. A genotype can be decomposed into units of heredity called genes. The different set of values that genes can store, also termed as *alleles*, are taken from a generic alphabet. The position of each gene in the chromosome is called *locus*. A set of robust operators are capable of exploring the search space in order to guide the current population towards more efficient areas of the search space. These operators, named selection, reproduction, crossover and mutation, are tools with a problem-specific implementation (Datta et al., 2007) and their particular structure influences the way the information is modelled. By the application of these operators, which try to emulate those found in nature, the strategy simulates the evolution of a population of solutions.

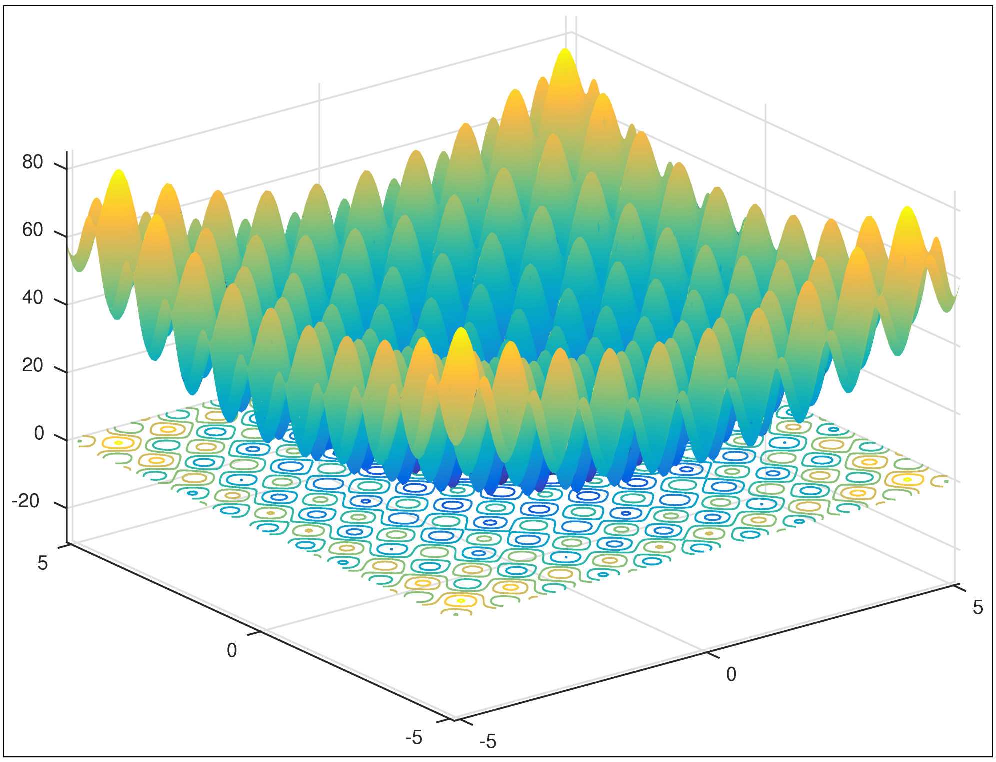



Figure 2.3: Example of a landscape commonly used to test the efficiency of evolutionary algorithms for two variables. The *Rastrigin* function first proposed by Rastrigin (1974), is a non-convex, non-linear multimodal function typically used as a performance test due to its capability of representing a large search space with numerous local minima.

The design of the algorithm should balance two factors: exploration and exploitation. If exploration is over-developed, the search process becomes inefficient; if the search is mainly focused on exploitation, it can easily reach *premature convergence*, losing diversity and getting trapped in a local optimum. This balance is pursued by controlling different search factors, such as selection pressure or operator design.

EAs are appropriate for large and high-dimensional solution spaces with both a continuous or a discrete nature, and particularly good in cases where prior knowledge of the space of solutions is limited. EAs have also been shown to be a flexible and powerful tool to solve non-continuous, non-linear, non-convex, multi-modal, non-separable, high-dimensional and non-differentiable objective functions with single or multiple objectives (Anagnostopoulos and Mamanis, 2010; Davis, 1991; Goldberg, 1989). In this regard, EAs are generally superior to other search techniques, which are limited by the continuity, differentiability, and unimodality of the evaluated functions.

There are multiple main branches of EAs that have been developed in parallel, largely independently one from each other. A list of the most important classes:

- Evolutionary Strategy (ES) (Rechenberg, 1973). This method, developed in Germany, was specially orientated to solve optimisation problems. It can be characterised by the use of string representations and the enhancement of both mutation and recombination operators. The selection operator is deterministic and the amount of population among generations is not constant. Typically ESs do not include a population of solutions but a single individual which evolves through successive application of the mutation operator and guided by the fitness function.

- Evolutionary Programming (EP) (Fogel et al., 1966). The intended area of applicability was the study of the evolution of finite state machines. The main characteristics of this approach is the emphasis on the mutator operator, the complete lack of recombination and a defined selection operator of probabilistic nature. The mutation operators were used to alter the states that were being evolved for the given tasks.

  The use of this strategy was insignificant for many years until its rebirth in the 1990s (Fogel, 1992) with much similarity to evolution strategies. EP, along with ESs are normally focused on continuous optimisation problems.

- GA (Holland, 1975). GA may be considered the most recognised form of evolutionary algorithms. The approach emphasises the effect of selection, recombination and mutation on the genotype, highlighting recombination (crossover) over mutation as the central operator. Mutation is only applied with a very small probability, typically less than 1%. The selection operator is probabilistic and it is called *proportional selection* and in its traditional representation uses a fixed-length binary chromosome. GA is considered more specifically suited to discrete and combinatorial optimisation problems.

- Genetic Programming (GP) (Koza, 1990, 1992). GP is another important sub-type of evolutionary computing that represents a major change in the way that solutions are encoded. The approach was explicitly aimed at constructing an evolutionary methodology for automatic programming which allows the design of the structure of the computing approach, instead of as a parameter optimisation method. Programs are normally very specific and small in size.

  The most widely used form is the Koza trees-based GP (Koza, 1991), also known as syntax trees or parse trees. Areas of applicability include the design of circuits, programmatic expressions and symbolic regression to fit a mathematical expression to data. A major problem in the GP paradigm is what is called 'bloat', which refers to the tendency for trees to grow large during evolution in a quadratic complexity that may lead to inefficient and uninterpretable programs (Langdon, 2000). In this regard, Luke and Panait (2006) analyse a series of bloat control methods for GP.

### 2.7.1 Single-objective Approaches

Starting from the initial set of individuals using a chromosome-like data structure, which are normally generated randomly, the algorithm tries to improve the population of solutions by evolving them over successive iterations based on the values of their fitness function. Each iteration corresponds to a new *generation*. To move from one generation to another, a new intermediate population, which forms the mating pool, is created taking information from the current population. By applying the crossover and mutation operators, reproduction is performed. In this exchange of information, the sources are called *parent*s and the new individuals are their *offspring*.

The number of parents can be a fixed number or can vary. Once the new offspring is created, a replacement mechanism defines the evolution of the population, deciding on the survivors for the next generation.

As the process continues, a sequence of generations evolves, increasing the average fitness of the chromosomes until any stopping criteria are met. Once the algorithm is halted, the final returned chromosome is the individual with the best fitness of the population.

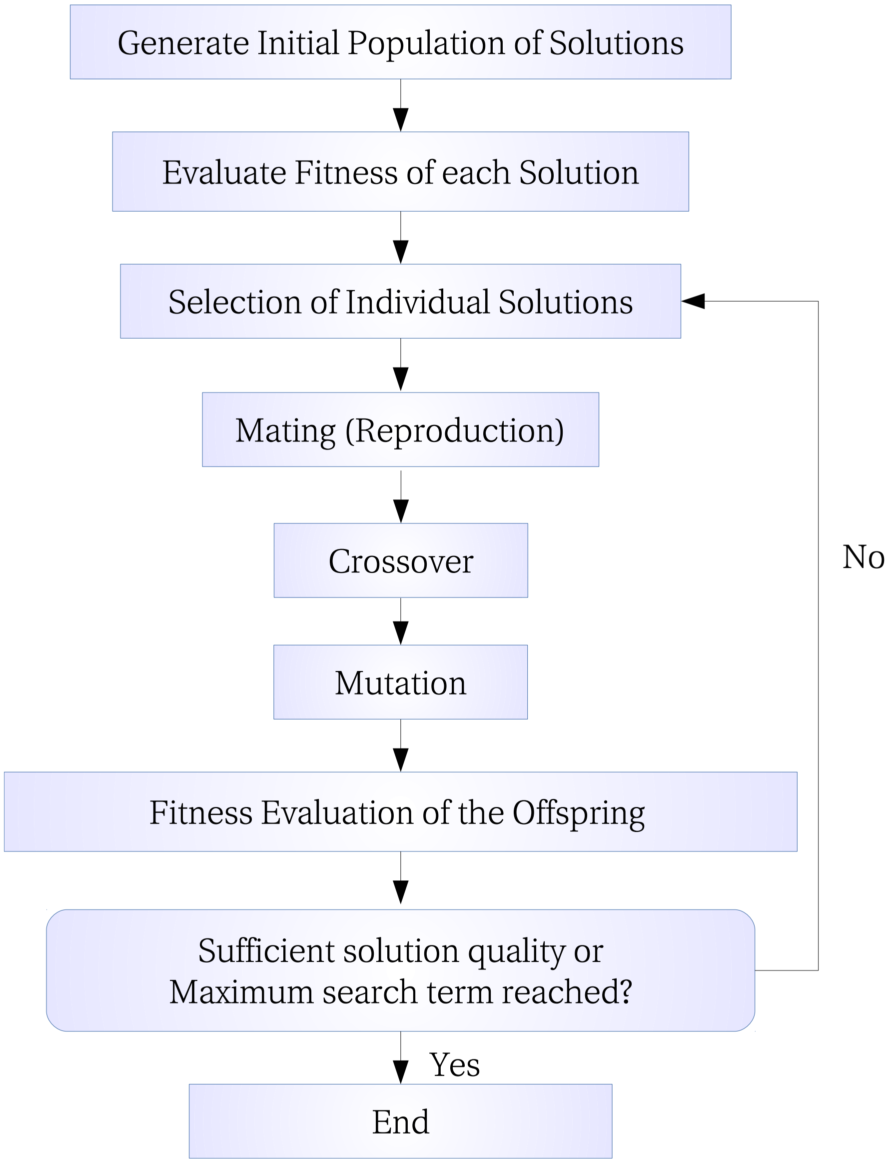



Figure 2.4: Basic structure of the evolution flow corresponding to a generic EA algorithm.

#### 2.7.1.1 Population

One of the most important differences between EA and other traditional methods is that EA manages a set of candidate solutions that is called a *population*. This population, created randomly or by a constructive algorithm, allows the exploration of promising areas of high performance of the search space. Population size has a deep impact on the performance of the algorithm. It is a parameter that can be maintained constant or can vary during the execution of the algorithm. A large population may result in a very slow rate of convergence, but at the same time contributes to the avoidance of suboptimal solutions. Small populations may easily lead to premature convergence but also have a quick rate of convergence.

Each of these solutions is called an *individual* or a *chromosome*, and can be seen as an abstraction from a biological DNA chromosome, consisting of a set of elements named *genes*. In biological terms, the representation of a given solution as a codification of a solution is called the *genotype* and in the search space *phenotype*. EAs offer a very flexible range of ways to represent solutions. Each particular representation, that is problem-dependent, is referred to as the *encoding*. The encoding can use different elements such as bits, numerical values, trees, arrays, lists or any other more advanced objects. The type of operators that can be applied to the chromosome depends on the type of encoding selected.

The significance of each element in the chromosome can depend on its value, its value and position like a binary encoding and only in its position. Since the traditional encoding for GA was binary encoding (a string of 0's and 1's), this is one of the most common used. This type of encoding is simple and allows the use of multiple implementation of crossover and mutation operators. When only the position in a sequence is important for the solution, the encoding is called *permutation* and it is used mainly in ordering problems. Permutation encoding permits the use of inversion and crossover versions that, in general, are complicated to implement. Tree encoding is the typical choice for evolving expressions in genetic programming.

Depending on the encoding and the type of operators applied, infeasible solutions can be present in the population. An infeasible solution is a solution that cannot be realised as an actual solution to the problem at hand. Sometimes, instead of dismissing these kinds of solutions, a penalty function is designed and applied to punish a solution that is not feasible. Other methods use repair algorithms to convert the infeasible solutions into feasible ones that occupy the feasible space $\mathbb{F}$. Repair is an ad-hoc mechanism that can be seen as a sub-state of the evolution process. The encoding structure is a critical decision that has multiple consequences on the performance of the algorithm, since a bad chromosome representation can increase the size of the search space or slow down the algorithm if too many repair operators are needed to ensure the chromosome is valid.

A review of methods to handle different constraint techniques can be found in Michalewicz (1995). Yeniay (2005) presented a review about multiple penalisation functions and Miettinen et al. (2003) compared some of the penalty-based techniques.

#### 2.7.1.2 Fitness

Every possible individual has an associated objective function value, calculated by the fitness function, defined to measure the optimality of a candidate solution through the adaptation to its environment. By inheriting positive genetic information between one generation to the other, the average quality of the population improves with time. Without this element, the evolutionary process would not be capable of progressing towards better areas of the search space. The posterior ranking of the population according to its fitness value is used to assess the suitability of the chromosomes, to implement the elitist mating and to select the final best solution. Different kinds of metrics can be used in the definition of this function according to which aspects are required to be emphasised within the system.

#### 2.7.1.3 Selection

The selection operator chooses individual solutions from the population using the fitness value, to include them in the mating pool as parents of the new offspring. If elitism is added to the selection strategy, the process tries to emphasise better solutions within the population by prioritising individuals with the better fitness over the rest. Without selection pressure, the search becomes more random and cannot emphasise more promising areas of the search space (Li et al., 1992). However, the overuse of selection pressure can provoke a narrowing of the population diversity (Whitley, 1989) which can lead to premature convergence which hinders the efficiency of the algorithm. On the contrary, too little pressure causes random search behaviour (Eshelman and Schaffer, 1991; Goldberg, 1989).

To determine which individuals can be more appropriate to give information to the next generation, a *neighbourhood function* is associated with the population $P$ in such a way that it can be formalised as follows:

$$N : P \to \rho(P) \tag{2.3}$$

where $\rho(P)$ is the set of the subsets of $P$. Each individual $e$ has a subset $N(e)$ of $P$ called *neighbourhood* that is the different individuals that can be created in one generation with the application of the available operators. These relationships between parental individuals are symmetric: $e_1 \in N(e_2)$ implies $e_2 \in N(e_1)$.

A variety range of selection schemes has been developed, each characterised in different ways. A review of the most important methods can be found in Bäck et al. (1997). Some of the methods are characterised by the following features:

**Tournament Selection** Tournament selection is a robust selection mechanism which is easy to implement and commonly used in evolutionary algorithms. In each generation a number $N$ of chromosomes are chosen uniformly at random, and the best of these is placed in the intermediate population. $N$ denotes the tournament size of the selection method. In expectation, this produces a linear ranking with a bias related to the tournament size toward the best individual.

The number of chromosomes selected in the tournament has a direct influence on both the diversity and the convergence rate. It is also defines the selection pressure. The bigger the tournament is, the faster the convergence is achieved. However, this comes with a premature convergence risk which could lead the algorithm towards local minima. The flexibility of this selection scheme lies in its capability of adjusting to different domains and problems by varying this tournament factor. This selection scheme introduces noise due to the random selection of individual solutions for the tournament, allowing that average quality or poor solutions may have children.

**Roulette Wheel** In this scheme, each individual of the population has a probability of being chosen, proportional to their relative fitness. This probability is calculated as a proportion of the sum of the fitnesses of the entire population. Then, better individuals have higher chances to be selected and reproduce. The roulette wheel has $N$ equally spaced pointers that are going to simultaneously pick the same number of individuals. The probability $p_i$ that the i-th member being selected is the following:

$$p_i = \frac{f_i}{\sum_{k=1}^{M} f_k} \tag{2.4}$$

where $f_i$ represents the fitness of the i-th individual and $M$ is the total number of individuals in the population. The resulting selection is unbiased (Baker, 1987).

**Stochastic Universal Sampling (SUS)** SUS is a non-bias, sequential and single phase method developed by Baker (1987) that selects explicitly each individual a number of times proportionally to the expected value of its fitness. The method creates $N$ equally spaced pointers, instead of a single one used in the roulette wheel method, where $N$ is the number of individuals to be selected in each generation. Afterwards, the population is shuffled randomly and a single random number $k$ is selected by generating a uniform random number in $[0, 1/N]$. The final $N$ individuals are selected by generating $N$ pointers, starting in $k$ and spaced by $1/N$, and choosing the individuals whose fitness falls on the positions of the pointers.

#### 2.7.1.4 Crossover

The crossover operator is a recombination mechanism inspired by the recombination of partial structured genetic material that occurs in nature to form the genotype of an offspring during reproduction. This exchange of characteristics is used to explore new parts of the search space with the intent of improve the fitness value of the next generation. In the mating process different members of the population (usually two), the parents of the new generation, combine the most desirable set of features in normally one new offspring. Crossover is applied according to a probability value $p_c$, called the *crossover probability*. If this operator is not included, then the variable $p_c$ is equal to zero.

One type of crossover commonly used is *one-point crossover*, see Fig. 2.5, where a position in the range $(0, N-1)$ is selected with a homogeneous probability, where $N$ is the length of the chromosome. The new offspring chromosome is formed by using the characters from the first section of the parent until the point indicated by the position of the crossover, filling the rest with the characters of the second parent. Other possible methods include two-point and multi-point crossover (N-Point crossover), uniform crossover, cut-splice crossover and three parent crossover.

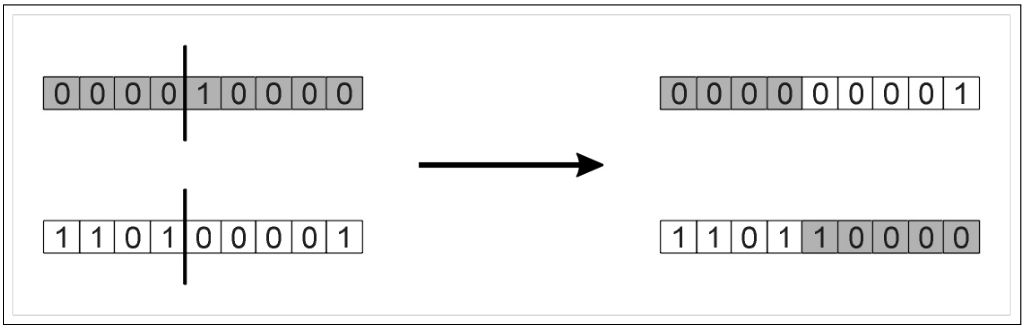



Figure 2.5: One-point crossover operation

If the crossover rate is too high then highly promising areas of the search space are discarded faster than the selection mechanism can produce improvements. However if the crossover rate is too low, the search may be stagnated due to the lower exploration rate (Grefenstette, 1986).

#### 2.7.1.5 Mutation

The mutation mechanism works by slightly modifying the structure of a chromosome and it is used to maintain the diversity and prevent premature convergence to local optima. The operator introduces stochastic perturbation in the new offspring by adding into them completely new information. This perturbation reintroduces back genetic diversity into the population and allows the presence of chromosomes that belong to unexplored areas of the search space. The mechanism also contributes to preventing convergence of the solution to a local optima.

Mutation is generally applied at a gene level. The random mutation selects randomly a gene within the chromosome with a probability $p_m$ to alter it to a new value selected from the whole range of possible values. On the contrary, the non-uniform mutation changes the values from a predefined distribution. Mutation rates are commonly very small and depend on the length of the chromosome. Consequently, after applying the operator the mutated chromosome will not be very different from the original one.

#### 2.7.1.6 Replacement & Stopping Criteria

After the offspring is created, the old population can be deleted to make space for the new generation. In the case the whole population is changed in each individual generation, the replacement is called *generational replacement*. Otherwise if typically at most one new solution is added to the population, and one is removed then the process is called *incremental/steady state* (Whitley and Kauth, 1988). In the case of a steady-state, for a population size of $\eta$, if $\lambda$ is the number of individuals replaced by new offspring, the *population gap* denoted by $\frac{\lambda}{\eta}$ defines the percentage of replacement. Whitley and Kauth (1988) reported higher convergence speed in the incremental method compared with the generational method.

The replacement strategy can follow different criteria like fitness-based (selecting the worst), age-based (the oldest) or simply choosing a random individual. The replacement condition indicates whether or not the replacement will be finally carried out. A common way to perform the substitution is to replace the parent only if the new individual has a higher fitness (Wakunda and Zell, 2000). In Goldberg and Deb (1991) is suggested to substitute the worst in the population to achieve a higher selective pressure.

The condition of termination can be defined following different criteria such as:

- *Number of generations:* The algorithm is finished when a predefined maximum number of generations is reached.

- *Elapsed time:* The halting condition is triggered when a concrete amount of time has passed from the beginning+ of the algorithm.

- *No change in fitness:* The genetic process will end if no change to the population's best fitness or any of the fitnesses of the population occurs for a specified number of generations.

### 2.7.2 Multi-objective Problems

Individuals, organisations and communities have different objectives and aspirations, with different views about the importance of factors such as ecological conservation, social rights and economy and the way of pursuing these goals. In order to maximise their point of view very commonly these stakeholders may organise themselves in pressure groups to influence decision-makers.

Therefore, many real problems can be more naturally defined by multiple objectives, usually conflicting with each other, that have to be considered simultaneously. Decision-makers, policy-maker and planners are willing to find a perfect solution that simultaneously optimises every defined objective, however often this is an impossible task. The procedure should be defined as the capability of managing a varied range of objectives together in order to find an acceptable trade-off among them. Nevertheless this is usually rather more complex than dealing exclusively with single and independent goals.

In a single objective problem the best single solution is the goal, however in a multi-objective problem frequently there is not a single best solution and, in the absence of further information, it is not possible to assess which one is better than the others. Therefore, in this case the goal is to select one solution from a set of promising ones. A promising solution should be good enough for all the objectives into consideration. The set of all feasible solutions $\mathbb{F}$ is referred to as the Pareto optimal set (Tušar and Filipič, 2015). In each Pareto optimal set, the solutions that cannot be improved with respect to any objective without worsening at least one of the other objectives are called the Pareto front or frontier (Fig. 2.6). For many problems, the number of Pareto optimal solutions is huge and the search of solution optimality is computationally intractable. Therefore, a practical approach consists of finding the best-known Pareto set that best approximate the Pareto front in a way that there should be evenly distributed over the Pareto front in order to show a real perspective of trade-offs to the decision-makers.

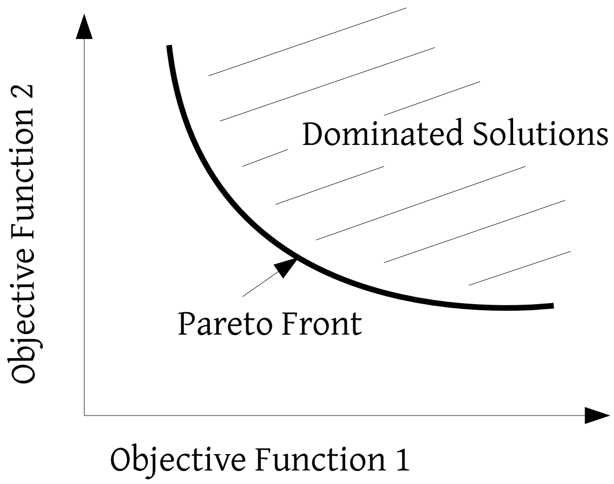



Figure 2.6: Pareto set/domain in a bi-objective space

Formally, the dominance relationship for a maximisation problem can be defined as follows: if $f_1, f_2, \ldots, f_k \in IR^k$ are the objective functions of a given problem, then it can be said that solution $x_i$ dominates $x_j$ which is denoted as $x_i \prec x_j$ if the following conditions are met:

$$f_m(x_i) \geqslant f_m(x_j) \quad \forall m \in \{0, 1, 2, \ldots, k\} \tag{2.5}$$

and

$$f_m(x_i) > f_m(x_j) \quad \exists m \in \{0, 1, 2, \ldots, k\} \tag{2.6}$$

where $k$ is the total number of solutions. These constraints means that $x_i$ should not be worse than $x_j$ in all objectives and it should be better than $x_j$ in at least one objective. Based on this concept of dominance, if $X \subseteq IR^k$ is a set of vectors, then the Pareto set $X^*$ of $X$ can be defined as follows:

$$X^* = \{x_j \in X \mid \nexists x_i \in X : x_i \prec x_j\} \tag{2.7}$$

The goal of the algorithm is to minimise the distance to the Pareto-optimal front, maximising the diversity of the generated solutions in terms of parameter values.

Typically classical optimisation methods have dealt with these kinds of problems by converting the multiple objective problem into a single objective one via methods such as utility theory or a weighted sum. The main advantage of this type of strategy is that normally the implementation is straightforward. For instance, in the weighted sum approach, it is required to associate different weights $w_i$ to each normalised objective function $z^{'i}(x)$ in order to combine them into a single scalar objective function as follows:

$$\min z = w_1 z^{'1}(x) + w_2 z^{'2}(x) + \ldots + w_k z^{'k}(x) \tag{2.8}$$

where $z^{'i}(x)$ represents the normalised objective function $z^i(x)$, $\sum w_i = 1$ and $k$ is the number of objectives considered. In the case that the particular problem requires the algorithm to return multiple solutions, the function has to be solved one time per solution needed using different weight combinations.

Classical methods following this approach include Linear Programming (LP) (Schrijver, 1998) which has been used, for example, to optimise different land-uses (Maoh and Kanaroglou, 2009). For further information, the survey about multiple objective integer programming of Figueira et al. (2005) can be mentioned, where different scalarisation techniques are reviewed. Other studies generate a unique objective underpinning the use of metaheuristics. In this regard, Hajela and Lin (1992) applied weighting-based Genetic Algorithms that focus on the design of structural and mechanical systems. Additionally, random weighted Genetic Algorithms were investigated in Murata and Ishibuchi (1995) where the multiple objective functions were randomly specified for each selection.

The major problem of such approaches is that they critically depend on the way the weights are defined (Masoomi et al., 2013). Besides, small perturbations in the weights may produce rather different solutions. Furthermore, not all Pareto-optimal solutions can be investigated. Optimality in non-convex Pareto front solutions cannot be calculated by minimising linear combinations of objectives (Cao et al., 2011).

#### 2.7.2.1 Multi-Objective Evolutionary Algorithms

As previously mentioned, in recent years, methods have been developed for problems in which trade offs need to be found among multiple objectives simultaneously. Using Multi-Objective Evolutionary Algorithms (MOEAs) to deal with such problems has been extensively studied by multiple authors (Fonseca and Fleming, 1995; Deb, 2001). The main advantage of EAs to address multi-objective problems is that it is a population-based algorithm that deals with a set of individual solutions that can search for many non-inferior solutions in parallel. This confers the ability to find multiple Pareto-optimal solutions in one single simulation run and makes EAs very attractive for solving multi-objective problems.

Since the appearance of the first multi-objective Genetic Algorithm study, conducted by Schaffer (1985) and introduced with the name of the *vector evaluated GA*, a number of different EAs methods were suggested to solve multi-objective optimisation problems. The most significant approaches include Non-dominated Sorting Genetic Algorithm (NSGA) created by Srinivas and Deb (1994), Multiple Objective Genetic Algorithms (MOGAs) developed by Fonseca and Fleming (1993) and the Niched Pareto Genetic Algorithm (NPGA) by Horn et al. (1994). These first approaches suggested the necessity of supporting the algorithm with additional operators in order to convert a single objective EA into a MOEA. Firstly, the measurement of the fitness within the population allows sorting each individual solution by the use of their domination value and secondly preservation of the diversity among solutions of the same non-dominated group it is used a niching strategy.

The next advance in MOEAs was the introduction of elitism techniques to facilitate convergence (Zitzler et al., 2000). Implementing elitism mechanisms in MOEAs is not as straightforward as in single objective optimisation. The main reason behind that is the large number of possible elitist solutions. However, its inclusion generally allows the algorithms to outperform their non-elitist counterparts (Zitzler and Thiele, 1999; Van Veldhuizen and Lamont, 2000). This elitist mechanism can be implemented using two different conceptions (Jensen, 2003): the maintenance of elitist solutions can be done in the same population or in an external secondary data structure, where they are kept until the posterior reintroduction.

From this elitist perspective, a new list of MOEAs was developed. Mentionable are the following well-studied elitist MOEAs: the Strength Pareto Evolutionary Algorithm (SPEA) developed by Zitzler and Thiele (1998) with an improved version called Strength Pareto Evolutionary Algorithm II (SPEA2) (Zitzler et al., 2001), the PAES introduced by Knowles and Corne (1999), the improved version of NSGA, called NSGA-II, proposed by Deb et al. (2000), the Micro-Genetic Algorithm ($\mu$GA) developed by Coello and Pulido (2001) and the Rank-Density based Genetic Algorithm (RDGA) presented by Lu and Yen (2003). For more detail refer to the numerous surveys related to MOEA (Xiujuan and ZhongKe, 2004; Coello, 1999, 2000).

#### 2.7.2.2 Pareto Archived Evolution Strategy

PAES (Knowles and Corne, 1999) is a simple MOEA algorithm that combines the use of local search techniques with a hill-climbing and random mutation strategy. The calculation of the quality of new candidate solutions is supported by means of the information provided towards a set of diverse non-dominated population solutions. In its original variant called $(1 + 1)$-evolution strategy, the algorithm uses a unique-parent and a single-offspring that are compared in each iteration. The creation of the offspring is generated by the use of binary strings and a unique bitwise mutation operator which is compared with its unique parent. This latter factor differentiates this approach from other MOEAs that maintain a population of solutions. The strategy uses an archive of previously visited non-dominated solutions to estimate the dominance rating of the new solution. A maximum size of the archive refers to the desired number of final solutions. Based on that structure, the algorithm is capable of distinguishing between good and bad quality solutions.

The operation of the Algorithm 2.1 goes as follows. The algorithm starts with the generation of an initial individual and its evaluation using the multiobjective fitness function. In each iteration this new candidate solution and a mutated offspring copy must be compared for dominance. If the offspring dominates the parent, then it is selected as the new parent. If the parent dominates the offspring, the offspring is discarded and the search of a new mutated offspring is performed. Finally, the information stored in the archive is used to compare them when both solutions are mutually non-dominating. If the offspring dominates any member of the archive, then this offspring is accepted as the new parent and all the dominated solutions are removed from the archive.

The pseudocode of the implementation of the PAES is depicted in Algorithm 2.1.

**Algorithm 2.1** PAES algorithm

```
global variables
    Max_iter
end global variables
local variables
Int_sol
Current_sol
New_sol
end local variables
procedure PAES
    Generate Int_sol, repair it at set it as Current_sol
    Evaluate fitness values of the Current_sol
    Add Current_sol to archive
    for i = 1 To Max_iter  do
        Randomly select one scenario for neighbourhood operation;
        Randomly select one case between (swap, reversion, insertion);
        Generate New_sol by neighbourhood operation in Current_sol, then repair it;
        Evaluate fitness values of the New_sol
        if New_sol dominates Current_sol then
            Set New_sol as Current_sol
            Update archive
        else
            if Current_sol dominates New_sol then
                Discard New_sol
            else            ▷ Current_sol and New_sol do not dominate each other
                Update archive using New_sol
                Select solution in the lesser crowded region;
            end if
        end if
    end for
    return Non-dominated solution
end procedure
```

In Algorithm 2.1, ▷ represents comments in the code, Int_sol is the generation of the initial solution, Current_sol is the solution that is currently being evolved, New_sol is the offspring solution generated by applying a mutator operator over the

*Current_sol* and finally Max_iter is the maximum number of generations that the algorithm is able to run. The time complexity of the algorithm is $\mathcal{O}(an)$, where $n$ is the number of generations that the solution is evolved and $a$ the number of solutions in the archive (Knowles and Corne, 2000).

However, in the cases where the offspring does not dominate any member of the archive and if the archive could exceed its maximum size, both parent and offspring can be included into the archive if they add diversity to the problem. In order to check this factor, a metric is used which is focused on the nearness among solutions stored in the archive. If the offspring is located in a least crowded region of the current search space in comparison to the rest of the members of the archive, it is accepted as a parent and a copy is added to the archive. The crowding factor is calculated recursively by dividing the entire search space into $d^n$ equal-sized hypercubes, where $d$ is the depth parameter and $n$ is the number of decision variables involved in the problem. A solution $A$ is located in a less crowded area of the search space than $B$ if $A's$ hypercube includes less individuals than the hypercube of $B$. The algorithm continues until a predefined and fixed number of iterations is reached.

In general terms, performance comparison among MOEA algorithms is a very complicated task (Knowles and Corne, 2002). However, PAES performance seems to be better in problems with a non-uniform density of the search space (Knowles and Corne, 1999).

## 2.8 Location-Allocation Problems

One tool widely recognised in the field of land-use change is spatial optimisation (Church, 2002). Models could be aimed at optimising the topological layout of a kind of resource (industrial plants, warehouses, health centres, wells, etc.) within a set of finite geographical candidate sites, to serve a set of spatially distributed customers. The final arrangement of facilities should meet a set of long-term goals that can cover concepts like ensuring proper accessibility to a type of service, maximising the coverage of the customers, or minimising the number of facilities, the associated costs or investment. The problem can be also formulated to anticipate the needs of a certain good in the future. In all of these scenarios concepts like distances, times or costs between customers and facilities are measured by the use of a given metric.

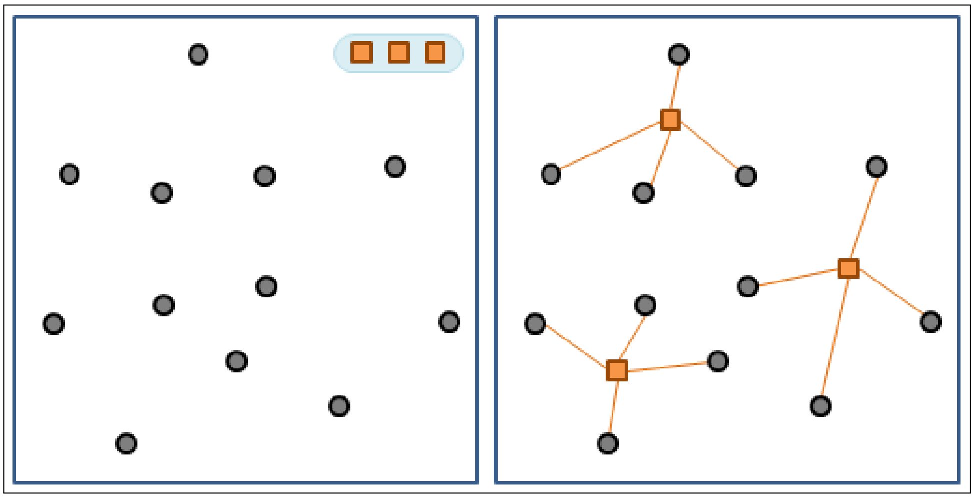


Figure 2.7: Example of a Location-Allocation problem where a set of three facilities are located to satisfy the necessities or entities represented as a blue circles.

Such problems were first introduced by Cooper (1963) with the name of LA problems. Commonly the allocation of such facilities implies large fixed investments and time-consuming activities associated with the opening, relocating and closing execution of a real estate transaction where in some cases, this operation also implies irreversibility (Antoni, 2002; Irwin et al., 2006). Consequently, allocation decisions should account for anticipated future conditions of the system under consideration. However, these future conditions are typically uncertain, and forecasts are frequently unreliable and subject to revision, so to produce quality predictions can be a very difficult task.

One way of subdividing the broad spectrum of location model is considering its space of representation(Daskin, 2008). Regarding this criterion LA problems can be divided into four categories, see Fig. 2.8. The easiest version is the analytic models, where the demand is homogeneously distributed and facilities can be placed anywhere. Typically these problems are solved using very simple techniques (Daganzo, 2005) like iterative LA algorithm, minima location model or median methods.

In the continuous version, the demand is not homogeneous any more. Instead it is only available at $n$ discrete defined points. The classical *Weber* (Weber, 1909) problem can be included in this class. The problem is to find the location of a unique facility to minimise the total distance between the facility and the customers. This problem can be solved by using simple iterative numerical procedures (Hamacher and Drezner, 2002) like the *Weiszfeld* algorithm (Weiszfeld, 1937).

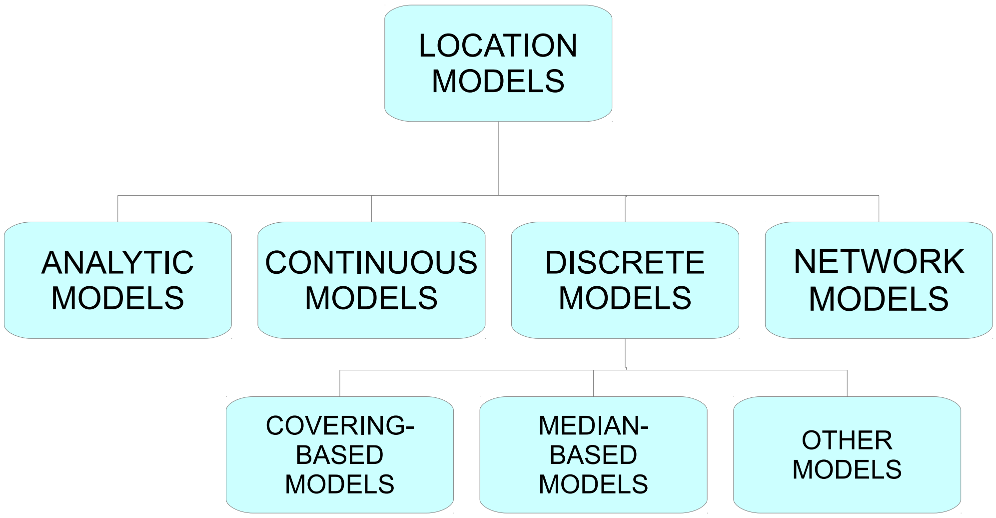



Figure 2.8: Taxonomy of location models based on the space in which LA problems are modelled.

In the network variant, facilities normally are located at the nodes of a network. When the demand rises, new locations can be found anywhere in the network. The main focus of this type of problem is to find polynomial time algorithms. Tansel et al. (1983) can be consulted for a review on this topic. Finally, in the discrete models, Cartesian metrics like Euclidean (Gray, 1997) and Manhattan distance (Krause, 2012) or Minkowski distance (Shahid et al., 2009), a generalisation of both are used to measure the distance or costs between locations. Generally, the placement of facilities is also restricted to a finite set of candidate sites. According to the taxonomy of Daskin (2008), this type of problem can be sub-divided into covering-based, median-based, and other types that do not fall into either of these two categories like the case of the p-dispersion model (Kuby, 1987).

Covering based models are characterised by a critical coverage distance or time within which demands must be fulfilled adequately. Such models are typically used for emergency services like fire stations. Median-based models are focused on minimising the average distance between a demand node and the assigned facility. Such models are essentially used for minimising the total transport cost. Finally, there are cases that the model cannot be included into any of these categories like the $p$-dispersion model intended at locating undesirable facilities (Kuby, 1987).

One model in the median-based setting is called the $p$-median problem, which has been extensively studied (Daskin, 1997; Hamacher and Drezner, 2002; Revelle and Eiselt, 2005). In this case, the configuration can be illustrated as the selection of a certain number of facilities subject to the minimisation of the total distance or costs in order to fulfil the customers' demands. In spite of its simplicity, Kariv and Hakimi (1979) demonstrated that $p$-median on a general network is NP-hard. The problem is additionally characterised by features like a single-period planning horizon, a deterministic set of parameters in terms of demands and costs and the allocation of single type of facility. Normally, such problems always consider full-connected paths between each customer and facility with a transportation cost equal to the travel distance.

One key factor of this type of problem is the distance to a given facility. Close surroundings have a stronger influence in land-use than more remote surroundings. This spatial relationship was numerously confirmed by empirical analysis of neighbourhood structure (Verburg et al., 2004). In this regard, long travel distances result in higher travel costs. If this cost reaches a certain critical threshold, people are less likely to use the service, searching instead for alternatives if they exist. This factor gives rise to two kind of allocation problems.

Following previous characteristics, in the uncapacitated LA problem introduced by Scott (1971), the capacity of each facility is unconstrained. This means that the nearest facility can always satisfy an unlimited number of customers (Mirchandani and Francis, 1990; Revelle et al., 2008). However, in a capacitated LA problem, customers' demand can be higher than that supported by the most convenient facility and hence it is not necessary that users select the nearest facility (Sridharan, 1995).

However, to model more realistic situations many extensions to the definition of the basic facility location problem have to be considered to overcome these insufficient and unrealistic settings. The model can be extended by the inclusion of stochasticity in any of the elements of the model(Snyder, 2006). The lack of deterministic behaviour can affect future customer demands, costs, the capacity of a determined location or any other element that needs to be inferred by the prediction of the future conditions of the system which implies always a certain level of uncertainty. In this context, Durmaz et al. (2009) considered the problem of finding the expected distances between the capacitated facilities and the customers using stochastic customer locations and deterministic customer demands and Mousavi et al. (2013) uses a mathematical model for the capacitated multi-facility LA problem where customers locations and demands are determined probabilistically.

Another important extension was intended to relax the single-period planning feature and consider instead a planning horizon divided into several time periods (Melo et al., 2006). As such, multi-period location problems, also called dynamic location problems, have been proposed to face situations where plans over several time periods are required to abate the entailed financial burden of such projects. The uncapacitated version was studied in Chardaire et al. (1996) and Galvao and Santibanez-Gonzalez (1992), meanwhile examples of works focusing on multi-period in a capacitated context are Fong and Srinivasan (1986) and Lee and Luss (1987). Apart from that, Owen and Daskin (1998) provides a more realistic model by giving an overview on facility location where is considered both previous extensions, time and uncertainty in some parameters.

An additional and crucial aspects is when the problem needs to deal with the arrangement of multiple heterogeneous facilities which may play different roles according to the relationships created among them. These facilities can be considered individually, or can be grouped into some kind of inner hierarchy (Şahin and Süral, 2007). The latter problem is called Hierarchical Facility Location Problem (HFLP) which aims to determine the most efficient and effective location of a range of facilities in a way that facilities at a higher level can serve those at a lower level. In this context, efficiency is related to financial objectives and effectiveness is linked with the idea of accessibility. This hierarchical representation is the general case of health-care systems (Farahani et al., 2014).

LA problems have been successfully applied to multiple disciplines like the location of disaster recovery centres(Dekle et al., 2005), train terminals (Horner and Grubesic, 2001), fire stations (Badri et al., 1998), coffee supply chain depots (Villegas et al., 2006), park-and-ride sites (Faghri et al., 2002; Farhan and Murray, 2008) among others. On the other hand, Malczewski (1991) investigated the environmental aspect of public facility location problems, studying the susceptibility of pollution in the context of paediatric hospital sites.

### 2.8.1 Solving Location Allocation problems

Standalone LA problems have been tackled using various traditional optimisation techniques, especially linear, non-linear and integer programming (Moore and Revelle, 1982; Ligmann-Zielinska et al., 2005), and branch-and-bound methods (Daskin, 2011; Kuenne and Soland, 1972; Malczewski, 1999). However, exact solutions are not computationally tractable for real-world problems because they may require exponential time to find the optimal solution, even for small instances of the problem. Besides, the amount of constraints and multiple objectives involved can cause that worst-case instances of the problem being NP-Complete (Ibaraki and Katoh, 1988; Zhang, 2002). For more information about complexity, refers to Gary and Johnson (1979).

But even if an optimal solution was found in reasonable time, there is no guarantee that the best exact modelling solution is also the best solution to the underlying real world problem. The fixed mathematical model for an exact method may not encompass sufficiently the real necessities of the environment. Heuristics, nevertheless, are more flexible and capable of copying with more realistic constraints.

As such, research on these problems over the past 20 years has mainly focused on robust meta-heuristics such as EAs (Shariff et al., 2012), Simulated Annealing (SA) (Murray and Church, 1996) and Tabu Search (TS) (Brimberg and Mladenovic, 1996; Ohlemüller, 1997) amongst others. Studies comparing these approaches on different tasks have reported contradictory results (Bettinger et al., 2003; Pukkala et al., 2004). However, formulation, implementation and parameter settings can significantly affect the final performance of the algorithms, which could introduce bias in the analysis (Crowe and Nelson, 2003).

### 2.8.2 Green Space Allocation

One of the most common interests in the urban community is the dynamics of urban growth, which is linked with the relative distribution of different industrial, cultural, recreational and commercial facilities, which may play an important role in a given urban environment. Their location has a direct economical and social impact on quality-of-life issues and on the broad strategies that are necessary to establish for creating an efficient land-use management (Robinson et al., 2012).

Public central facilities such as parks, libraries, schools, sports and health centres are characterised by people have to travel to receive the service offered. A major piece of research is done in finding the best placement for these central facilities in order to provide a proper service to the entire set of potential customers at minimum cost. This optimisation process is, in general, very sensitive to the distribution of population density and to its level of accessibility.

In the discipline of urban ecology, urban parks are considered one of the most important components in urban sustainability (Munro et al., 1991). They provide a healthy environment and promote quality of life to the urban population (Chiesura, 2004). However, the provision of public open space is in rivalry to other urban land-uses, simply because it consumes physical space. Apart from that, other factors can constrain the proper provision of these sites like limited funds for maintenance, personnel and supporting facilities (Boone et al., 2009). In spite of the importance of the problem, and while much research has been focused on the allocation of a varied set of different facilities, the green space allocation problem, as such, has received very little attention so far, with just a few exceptions.

Sefair et al. (2012) proposes a multi-objective location model that, by interacting with a GIS system, is capable of determining which parcels of land should be transformed into new parks using five indices: spatial coverage, number of beneficiaries, accessibility, nearby facilities and cost. Zucca et al. (2008) conducted a multicriteria analysis to select the best metropolitan parcels of land for being designated as green areas over four candidates, considering several factors including economic, social, and environmental criteria. Tajibaeva et al. (2008) developed a discrete-space urban model to analyse the optimal size and location of open green areas. They also analysed the effect of these green areas on the spatial and density urban development and the evolution of property values. The study is limited to cases with ($5 \times 5$) neighbourhoods. Following the same logic, Quaas (2007) investigates efficient urban spatial structures using a close-monocentric urban economics model that includes environmental amenities from public open space. Neema and Ohgai (2010) created a multi-objective model to determine the best location of urban green areas using a non-constraint and continuous area of study, selecting the most convenient parcel independently of its availability. The study includes as criteria the population density, the air pollution, noisy areas and the under-provision of green spaces.

The model of Wu and Plantinga (2003) is a representative example of the use of a theoretical model which follows Alonso's assumptions to study the allocation of green spaces. Specifically, the model analyses several predefined arrangements of green areas in a city to study the consequences of a determined urban spatial structure policy like the appearance of leapfrog phenomena. The development of leapfrog is characterised by a scattered form of urbanisation which is interspersed with green areas. Related to that topic, Brown et al. (2004) evaluates the effectiveness of other important green area structures such as greenbelts that are located close to emergent developed areas.

### 2.8.3 Value of Open Spaces

Green areas are public goods without a direct market price. However, following an empirical approach and indirect methods, econometric tools including hedonic pricing and contingent valuation can be used to measure the value of green areas (Brander and Koetse, 2011).

Hedonic pricing is a valuation method used to measure the value of urban park, estimating the impact on prices of residential properties located in the proximity of a green area (Tyrväinen, 1997; Bengochea Morancho, 2003). The method requires the collection of observations on property values. Contingent Analysis is a method applied to the valuation of green areas (Riera, 1993; del Saz Salazar, 2000; del

Saz Salazar et al., 1999; Breffle et al., 1998) which asks directly to people how much they will consider to pay for the use or maintenance of a particular good. However, in the case of open areas, both models provide only a partial quantification of the entire amount of economic benefits. This kind of land supports many services that may not be perceived as a market value (Mahan et al., 2000; Brander and Koetse, 2011).

The technique demonstrates that urban market shows green externalities because prices of properties increases with the proximity of urban parks (Tyrväinen and Miettinen, 2000; Thorsnes, 2002; Bolitzer and Netusil, 2000; Lutzenhiser and Netusil, 2001) and wetlands (Mahan et al., 2000; Luttik, 2000; Acharya and Bennett, 2001). Other factors could also exert influence (Tyrväinen, 1997), including the size of the urban park, and whether or not the total number of open areas in the region is scarce. Moreover, prices can be also modified by market expectations related to the future uses of surrounded undeveloped land (Smith et al., 2002).

Regarding the quantification of this relationship, the studies show different figures. Brander reported that residences increased their values by 0.1% when they are located 10 metres closer to an open space (Brander and Koetse, 2011). Tyrvainen, on the other hand, shows an average decrement in market price of 5.9% per kilometre (Tyrväinen and Miettinen, 2000). In Mahan, it is stated that prices increase by $24.36 per acre and by $436.17 per 1000 feet (Mahan et al., 2000).

When size of the green space is included in the study, it seems that price per area drops with the larger parcel size, which means that there is a decrement in the marginal value of the size. Brander shows that 10% larger parcel implies 8% lower price per hectare (Brander and Koetse, 2011). However there is a general willingness to pay more for dwellings close to larger areas of open space (Smith and Osborne, 1996).

### 2.8.4 Location-Allocation problems with a temporal dimension

Extending a location-allocation problem to incorporate a temporal dimension, covering the allocation of facilities during a period of time, received multiple names in the literature including multistage allocation problem, dynamic facility location problem or multi-period location problem (Warszawski, 1973; Van Roy and Erlenkotter, 1982).

The problem introduced by Ballou (1968) was focused on locating and relocating a single warehouse over a period of years, and its solution was implemented by means of dynamic programming with backward recursion. Other earlier studies worthy of mention include the work of Erlenkotter (1981) who reviewed the use of other heuristics for deciding when to add a facility capacity expansion so that demands in all periods may be satisfied, and Roodman and Schwarz (1977) which aimed at the study of declining or shrinking markets. da Gama (2002) and Wang et al. (2003) additionally include a budget constraint which limits the number of possible opening and closing operations.

Different time horizons can be considered in this type of problem. The essential scenario deals with forecast horizons, where population should be served from the beginning of the planning period (Albareda-Sambola et al., 2009). However, commonly problems are based on the allocation of non-essential facilities, where the horizon is defined at the end of the period considered. In some problems the idea is to seek an $\epsilon$-optimal forecast horizon, since proving that a policy is strictly optimal for any finite duration planning period cannot be done efficiently (Daskin et al., 1992).

A concrete version of this problem deals with situations in which parameters change over time in a predictable way. However, when stochasticity is included into the system, the problem can be addressed from the perspective of a Sequential Decision Making Problem (SDMP) (Howard, 1960). Generally speaking, a discrete-time finite sequential decision process can be defined as follows: there is an environment which can be described as a state-space set $S$ with distinguished initial state $s_0$ and an action set $A$ where $S$ and $A$ are both finite. Each state $s \in S$ is dependent on the previous state of the system and the action $a \in A$ taken. The transition function $\delta$ controls how actions modify the state of its environment.

$$s_{t+1} = \delta(s_t, a) \tag{2.9}$$

We define a policy $P_i$ such that the mechanism in charge of selecting the next action is based on the current perception of the environment. This perception can be total or partial:

$$\begin{aligned} &P_i : S \rightarrow A \\ &P_i(s_t) = a_t \end{aligned} \tag{2.10}$$

In turn, the action $a$ influences its environment, causing a change of the current state. The process starts in the state $s_0$ and by means of the sequential application of the policy $P_i$, further actions are chosen. The final sequence of decisions, one for each period, will be called a policy. An optimal policy for either a finite or infinite horizon problem should minimise the value of all present and future costs.

Several approaches have been applied to solve sequential decision problems, including Decision Trees (DTs) (Garcia and Sabbadin, 2006; Jeantet et al., 2012), Influence Diagrams (IDs) (Guezguez et al., 2009) or more commonly Markov Decision Processes (MDPs) (Bellman, 1957). The family of MDP algorithms is considered the most traditional approach to solve such problems. By observing the different states of the system, a central controller which has access to complete state information, searches for an optimal strategy by means of tools like rewards and punishments (Puterman, 1994). If the process is controlled by multiple cooperating distributed agents, each with possibly different information about the state, the problem is called a Decentralized Partially Observable Markov Decision Process (DEC-POMDP) (Amato et al., 2013), which is a generalisation of Partially Observable Markov Decision Processes (POMDPs) (Sabbadin, 1999). The main drawback, in these cases, is that such strategies cannot be generally scaled to large problems, since to properly represent these kind of problems, the number of required states grows exponentially (Pineau et al., 2006; Smallwood and Sondik, 1973).

### 2.8.5 Green Space Planning

The main purpose of planning, in the context of urban and other land-use, is to improve the community's quality of life by creating a better social, economical and physical environment by means of allocating land for specific purposes. Green areas and parks located close to urban areas are patches of land composed by greenery and trees normally formed by non-native species (Byomkesh et al., 2010) that are designed to provide amenities to residents in the form of recreational benefits.

In order to maintain enough provision of this type of land-use, new disciplines arise. In this sense, urban sustainable development can be defined as the preservation and the enhancement of positive qualities of the city, avoiding sort-term measures in order to perpetuate the characteristics of the system (Throsby, 1995). This concept is conceived as a central key to solve world environmental and social problems on a global scale. In this regard, the research community working in this area needs to examine the interactions between ecological and social dynamics, which can be seen as the trade-off between over-exploitation of land for profit, and assuring nature's resilience against perturbations.

Open space planning is a kind of facility planning which concretely deals with the management of natural areas and the development of social and environmental services. A proper strategy for open space planning is not only important because of its aesthetical aspects, but also as a part of the promotion of urban sustainable development (Sanesi and Chiarello, 2006; Jim and Chen, 2006) and ecosystem services provision. In fast-growing urban areas, it is required to build additional public facilities to address increasing demand for services in areas of population growth, while the opposite process of closing of facilities may be needed in areas of population decline. However, for implementing an optimal locating process, the study and forecast of the demands and costs involved is necessary. Demand could fluctuate over time, requiring opening and closure of new facilities. Consequently, decisions on when, where and the quantity of land are not independent of each other (Manne, 1967).

Recently, green space planning, and the advances in our ability to predict future trends and patterns to guide decision making, have gained increased attention in the research landscape. From the planning perspective, open space management is a challenging task that should be analysed as a complex system (Boyd, 2008), due to different plausible policies and management objectives may lead to multiple future scenarios. Many factors contribute to the complexity of open space planning, including the inherent ill-structured nature of the problem, the existence of spatial dependencies, the presence of multiple conflicting objectives, non-linear relationships between decisions and their consequences and varied forms of uncertainty (Janssen and Rietveld, 1990).

In urban land-use allocation, the consequences of allocating a land-use type to a particular area are highly context dependent. For instance, the dynamics of mixed land-used parcels like residential and commercial are linked with employment location by the defined transport network (Deitz, 1998). In the concrete case of green spaces, spatial relationships between residential locations and adjacent natural environmental units create dependencies amongst activities in surrounding areas; in turn, these introduce new non-linearities (Stewart et al., 2004). Another example of these relationships is the case of the application of urban densification strategies, which creates complex relationships between urban development and the quality, as well as the quantity, of urban green space (Arnberger, 2012).

These types of complex dependencies and uncertainties often presented in objective functions, constraints and system predictions, may severely affect the robustness of the underlying allocation processes. This could result in sub-optimal or even infeasible green area network designs, with the outcome being a significantly reduced overall level of provision.

Another issue that complicates the provision of green areas is that currently there are no available adaptive planning tools that can realistically scale to current conservation scenarios (Xue et al., 2012). Hence, it is necessary to develop a type of sustainable planning that tries to recognise and anticipate how patches of land influence urban patterns, their dynamics and further consequences for the population that its distribution entails, since the availability of patches most probably change over time. Furthermore, globalisation contributes to the promotion of new management structures, privatisation processes, and prioritisation of commercial objectives that can be potentially harmful and restrict democratic practices linked with the use of green areas (Littke, 2015). These effects are caused by the application of design policies that excludes some collectives of people (Low et al., 2009) and also these facilities are frequently conceived as invisible assets (Mayor of London, 2009).

According to Savas (1978), the evaluation of a determined allocation planning specially for public services should be based on three primary criteria: effectiveness, efficiency, and equity. Effectiveness can be conceived as the level to which this resource allocation accomplish to meet the needs of the objective individuals. Efficiency can be defined by the ratio of outputs or costs to inputs or benefits given a particular resource allocation. Finally, equity may be depicted as synonymous with fairness.

Carr (1992) highlighted five characteristics that open spaces within cities must offer, namely comfort, relaxation, accessibility, public ownership, and freedom. In this regard, governments can adopt a wide range of interventionist mechanisms to restrict the ownership over the land and control its use, acting as a response to social requirements over gardens and parks to provide a set of services based on the proximity to potential users. Among these measures, local authorities can assume the ownership of the land and assign them partially or totally the function of urban green spaces like the case of Stockholm city. The capital of Sweden is world-renowned for its protective environmental and sustainable urban planning, possessing huge areas of urban green space (Passow, 1970; Littke, 2015).

In the concrete case of parks and green areas, a successful allocation strategy should be based firstly, on finding the amount of parcels which ensures enough space provision and secondly, on how efficiently these areas are distributed. Even if there is a lack of consensus about how to achieve these goals (Maruani and Amit-Cohen, 2007), a trade-off between the available budget and the selection of the most expensive areas with expected highest impact on population should be found. This would achieve, through the time period considered, a higher service reward for the current and future urban population. Budget constraints typically make it difficult to achieve complete coverage of the population for each period, and therefore a sub-optimal plan is inevitable, which needs to maximise the partial service coverage at minimum cost over the entire planning period.

To fulfil these requirements, planners should decide which parcels to select from a noticeable large number of candidates. The number of possible subsets may imply a vast number of choices. In addition, due to the presence of constraints and multiple criteria, like for example to provide multiple services to the society and the conservation of biological sources, the choice of multiobjective optimisation would seem a good strategy to follow.

Nowadays, to find an optimal or near-optimal solution is not the unique possible objective. With the increased involvement of stakeholders, optimisation models are more often used as tools to support design of robust decision policies, rather than a strategy to generate the best alternative. These decision support-systems require short response times from the algorithm in order to adapt its solutions with the inputs from the stakeholders.

In land use planning, due to the changing nature of the land dynamics, only short term predictions can be securely applied (Cheng, 2003). The accuracy of the projections resulting from the use of models and simulation techniques is expected to decrease over time due to far projections further beyond the period for which the model was validated. These projections not only encompass great amount of uncertainty (Pontius et al., 2004; Tattoni et al., 2011) but also they depend on stakeholder needs and variations in economic pressure and related legislation.

## 2.9 Evolutionary Algorithms in Urban Scenarios

### 2.9.1 Single-objective

EAs have proven to be an efficient and effective tool for solving different geographical problems (Xiao et al., 2007) which is capable of finding near-best solutions in a reasonable time. Numerous studies have selected this type of technique to investigate spatial land-use problems.

EA can provide to the policy-makers the capability of evaluating alternative land-use configurations (Stewart et al., 2004), calibrate parameter values and transition rules within a CA framework that studies spatio-temporal urban growth (Tang et al., 2007; García et al., 2013) and improve the accuracy of the model (Clarke-Lauer and Clarke, 2011; Li et al., 2013). For other land-uses, EA has been applied to analyse land use planning in the field of forest structure optimisation (Venema et al., 2005), in Porta et al. (2013) where a parallel high-performance EA is used to deal with planning problems with a huge number of possible land category combinations and in Lehmann et al. (2013) where an agricultural land-use problem was studied by the development of a bioeconomic whole-farm model to support farmer's decision-making under different climate and price scenarios.

### 2.9.2 Multiple-objective land-use planning

Land-use planning is by nature a multi-objective problem. In a real-life urban planning scenario, these objectives conflict with each other, and optimising a single objective without respect to the others can produce non-valid results. Besides, it is very difficult for planners to numerically quantify their relative weight or importance, especially in a dynamic scenario. This can produce the situation that the highest priority objectives at one given time may not be the most important ones in a relatively close future. Then, to achieve a set of trade-offs based on the relative importance of each objective against the rest, we must look at the use of Pareto-sets. One advantage of this would be that these sets are usually independent of the relative importance of objectives considered (that is, this relative importance does not need to be pre-specified), and suitable for complex applications such as land-use planning.

A number of traditional optimization techniques have been proposed for the computation of multiple land use problems (Wright et al., 1983). In the field of land-use change we can mention its use in forest management (Ducheyne et al., 2006), land use planning (Stewart et al., 2004), urban planning (Balling et al., 1999), resource allocation (Datta et al., 2008) or determining the most convenient location of undesirable facilities (Rakas et al., 2004). Cao et al. (2011) investigate the development of optimal land-use scenarios using three objective functions: minimising conversion costs and maximising the accessibility and the compatibility level between land uses by implementing NSGA-II. Cao et al. (2014) uses also a NSGA-II approach to calibrate a cellular automata used to understand rural–urban land conversion processes. Melachrinoudis et al. (1995) developed a multi-objective version of the capacitated multi-period LA problem for landfill facilities where garbage can be disposed. Yang et al. (2007) investigates a fuzzy version of a Genetic Algorithm to optimise the location of fire stations. The work proposed by Sefair et al. (2012) uses a lexicographic order of the evaluation criteria and a maximum deterioration of the objectives with higher priority to support decision-making process to city planners regarding the location of new urban parks in Bogotá.

Other powerful heuristics were also studied in this context. Ma et al. (2011) selects Particle Swarm Optimization (PSO) to optimize the suitability allocation of land uses minimising costs and (Semboloni, 2004) developed a Simulated Annealing (SA) method to optimize both residential and commercial facilities.

## 2.10 Conclusions

This chapter has introduced the background theory and concepts that underpin the research work of this thesis. Starting with a discussion of green spaces, urban planning and urban models, it was previously argued that cities tend towards decentralisation. This tendency is also reflected in the evolution of urban models over time, which has changed from early top-down approaches to more recent bottom-up strategies. One of these bottom-up strategies to model such dynamics is CA. This tool has the intrinsic ability to represent spatio-temporal complexity and, in consequence, it is commonly used for the representation of land-use phenomena. Additionally, ABM is a technique commonly combined with a CA to include heterogeneous entities which spatially interact at an aggregate level within the model. This element allows us to include human behaviour into the urban model. Additionally, this model is hybridised with the inclusion of the classical economical model of Alonso, to address how agents select the most convenient household to settle down. The canonical model of Alonso provides a fast and simple way of including different factors of preference for a theoretical implementation of population dynamics. In Chapter 3, the implementation of the urban model and its components using these techniques is described and discussed in detail.

The chapter continues with the presentation of the two central optimization problems addressed in this thesis: the location-allocation problem and a dynamical planning extension of it, which can be seen as a SDMPs under uncertainty. The first is defined by the spatial optimisation of a set of facilities, while the second is a discrete planning problem where the outcome is influenced by the uncertainty linked with the future dynamics of the system. All of these elements are illustrated with examples in the context of urban development.

Finally, the metaheuristic used to solve such problems, evolutionary algorithms, are described in general terms. The application of EA in planning problems under uncertainty is not common in the literature, since it requires further tools to be able to deal with the uncertainty. Hence, the application of this concrete technique to this area of research is innovative. This thesis also investigates a mechanism to aid EA to cope with the uncertainty using Monte Carlo sampling. The description of the developed method is depicted in Chapter 4 and applied in Chapter 5 and 6.

# Chapter 3

# Urban Growth Model

## 3.1 Introduction

In this chapter we describe a computational model of urban growth, built using a cellular automaton and an agent-based system, which is based on Alonso's widely adopted theoretical framework for urban growth and land economics (Alonso, 1964). The chosen modelling approach can be viewed as a hybrid between top-down and bottom-up strategies. The model will be the basis for simulating urban growth in the context of planning decisions, and therefore play an essential role in evaluating the consequences of those decisions. The model incorporates a real estate market, and a dynamic collection of agents who operate in that market, interacting and moving in time and space. These agents represent families who wish to settle down in the city. Assuming that such a model can be characterised as a complex self-adaptive system that is capable of simulating land-use dynamics, the model allows us to analyse urban growth, population density and environmental dynamics and the way they interact as complex, interconnected phenomena. In this context, urban development includes both new development and urban redevelopment.

## 3.2 Model Description

The proposed theoretical urban growth model was implemented as a computer program in Java & RS (North et al., 2005) for analysis and visualisation purposes.

The CA-ABM framework selected to model the city is a discrete-event system, used to simulate neighbourhood interactions to mimic land dynamics. These dynamics rule the growth behaviour of a city, the rural areas which surround it along with the distribution of the population living within it. The physical layout of the city is configured by a two-dimensional regular discrete lattice of $50 \times 50$ contiguous squared cells with i- and j-axes in an Euclidean space $\mathbb{R}^2$ arranged within a canonical mono/multicentric framework. Land transformations are ruled by a set of stochastic transition rules that are applied on a cell-by-cell basis. The type of neighbourhood selected $\Omega$ is made up by the eight adjacent cells that define the Moore neighbourhood, see Fig. 2.3a. The use of the Moore neighbourhood allows the integration of diagonal and perpendicular dynamics. The evolution of the city is ruled by an internal schedule with a determined time-horizon of finite duration. Time gap has no substantial meaning. Each single time span determines one iteration of the CA where all changes are applied synchronously.

Every cell of the lattice corresponds to the surface of a physical unit area of the landscape under consideration $S$. It is assumed for simplicity that for all dwellings, the land occupied $S$ is fixed to a positive constant $\bar{S}$. Each of these patches of land can be identified by its location $(i, j)$ and by a unique land-use class whose value is dynamic through time. This land-use type represents the predominant land-use at that location. The general classification of land types under consideration comprises urban, rural and protected areas which, in turn, can be subdivided into more specific types.

Each cell in the model represents a unit of the landscape identified by its location $(i, j)$ and by a unique land-use class whose value is dynamic through time.

Let $C$ be the set of all single parcels of land in the considered geographic area represented in the CA, then the more general land-use subdivisions $N_U^t, N_R^t, N_P^t \subseteq C$

can be summarised by the following set of cells: $N_U^t$ denotes all the urban parcels in the grid, $N_R^t$ groups the type of rural land units and $N_P^t$ represents the open spaces protected at a determined time step $t$. All subsets are always mutually disjoint, which can be formalised as:

$$c_k^t(i,j) \in \{0,1\} \tag{3.1}$$

where $k \in K$ is the land use type, $K$ is the set of all possible land uses states $\{U, R, P\}$. Then, at a determined time $t$ the variable cell $c_k^t(i,j)$ is equal to 1 if the land-use $k$ is present at location $(i,j)$ and 0 otherwise. Those values are dynamic through time and influence other internal characteristics of the cell, such as its price and its ecological value. The city and its hinterland are modelled in such a way that:

$$N_U^t + N_R^t + N_P^t = NC \quad \forall t = 0, 1, 2, \ldots, T \tag{3.2}$$

where $NC$ is a constant value that represents the total number of cells in the grid and $T$ is the maximum time horizon of the simulation.

Assuming that the city is located in a featureless and flat terrain, the model can depict an urban development with one or several CBDs (Fig. 3.1) where central cells tend to absorb most of the population. These cores represent not only employment centres, but also shopping points generated by consumer decisions, business interdependences and clustering of jobs.

Each urban cell can be populated by more than one family unit that, in turn, comprises an adult agent and its offspring. Then, every set of agents that compound a family are always located in a specific place $(i,j)$ within the boundaries of the city, at a certain distance $x$ to the edge of the urban area $(\tilde{\imath}, \tilde{\jmath})$. The model can be defined as an open-city framework with perfect mobility where flows of incoming endogenous population can be accommodated within the city. This positive flow is based on urban migration is attracted by potential urban employer offers where job opportunities are large in number and hence develop faster.

The model starts with a completely undeveloped grid, where every transformation of activity is new to the land parcel under consideration. A similar initial configuration

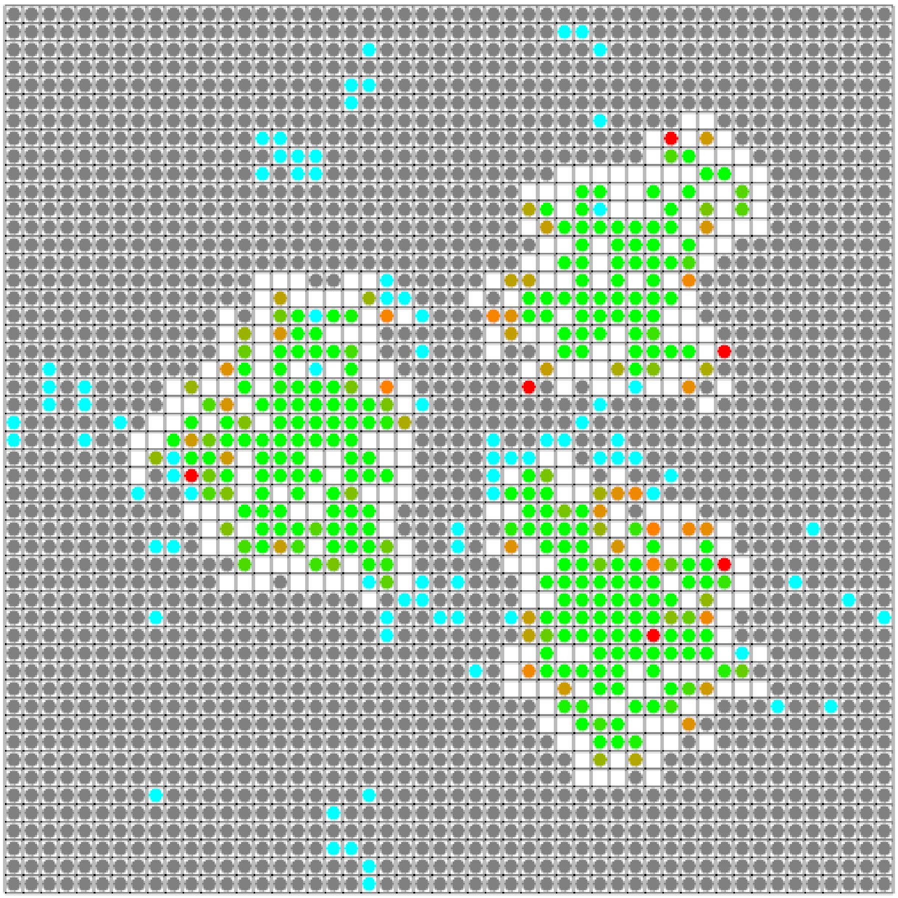


Figure 3.1: Simulation of a city in expansion. The grid layout is configured with three CBDs that grow in parallel. Green areas are represented in blue, urban developed land in green, new land under construction in white, land that was just transformed to an urban type, ready to be built in red and rural land in grey.

was used in previous models (Church, 2002, p. 10; Ogawa and Fujita, 1989), while other approaches start with an initial landscape arrangement (Nalle et al., 2002; Ligmann-Zielinska et al., 2005). The individual aggregated decisions and the defined transition rules combine to lead the city to stochastically expand its peri-urban boundaries through time until a determined time horizon is reached. By the use of its transition rules, the cellular automata is potentially able to change the state of every cell in each time step. These transition rules, that represent land spatial relationships within the urban development process, are conditioned by different probability values which introduce stochasticity into the system.

Following an extension of the classical canonical economic model of Alonso (1964), where multi-centricity and green externalities are included as model extensions, householders modelled by an ABM look for an economic competitive equilibrium between housing space and commuting costs. Green externalities are defined as the spatial relationships created between green areas and household prices (del Saz Salazar and García, 2007), which induce preferences for housing locations close to green amenities and, consequently, modifies urban patterns in the long term.

In conclusion, the dynamics of agents and cells allow the model to evolve among a set of predefined one-directional states at each time step. The individual aggregated decisions and the defined transition rules of the city to stochastically expands its peri-urban boundaries through time. The final topological patterns and the speed of the urban process depend mainly on the CA transition rules, the residential choices of the inhabitants and the location of green areas within the city, where people generally prefer to live (del Saz Salazar and García, 2007). From this perspective, the model is therefore dynamic in time and space, and each simulation run will yield a different result.

### 3.2.1 Extension from previous model

#### 3.2.1.1 Software Design

An in depth review and transformation of the class structure and component was performed to follow the standard Object-Oriented Programming (OOP) convention. This process included: access to member variables, also called fields, from input output methods instead of direct access; adding a system of exceptions, instead of managing errors by using boolean variables to avoid the abrupt finishing of the program in case of an unexpected failure; changing the scope of the member variables to *private* to enhance the level of encapsulation of the code and adding *static* behaviour to the system to allow the interchange of constant values and avoid redundancy.

Some composite classes were added to divide the code related to the common behaviour of multiple objects that belong to the same class. Concretely, a class that unifies all the cells in the grid called *Lattice* was added. Additionally other classes that manage only the part of the grid related to a specific land-use type were also implemented. These classes were *City* that groups all the urban cells in the grid and

*GreenArea*, which does the corresponding with the rural and protected cells. This strategy was followed in the case of the population. A corresponding *Demographic* class was implemented to gather the common behaviour from each agent of the grid. The accessibility among classes was restricted to enhance the composite classes act as interfaces to the rest of the system.

After the inclusion of these classes, relationships between classes that were not directly affected by an inheritance or composition relationship were minimised, using always the most general class to perform exchange of information (input and output). Connections among classes formed a graph that was almost fully-connected. Due to that, an effort was made to remove the non-necessary dependencies without modifying drastically the code.

An attempt to lighten the code was performed by removing redundant information from the system in different ways and reimplementing some critical functions. The preliminary code stored their neighbourhood in each object *cell*. This requires an additional eight cells with the corresponding constants and data structures. This approach is very demanding in terms of memory usage, requiring eight times more memory to store the complete lattice, since the neighbourhood defined in the system is composed of eight cells plus the centre cell. Then, if the size of the grid is $25 \times 25$ then, the number of cells constantly loaded in memory is $2500 \times 8$. The code was changed to calculate dynamically these cells each time the neighbourhood of a single cell is required. Since the area under consideration was enlarged from the previous piece of software, these kind of cleaning and restructuring operations were crucial to avoid the collapse the system at the end of the simulation, when more objects are created in the model.

#### 3.2.1.2 Discovery and management of errors in the code

The code was reviewed and several problems were fixed:

- A general review of the code that control the generation and release of agents in the system were done. Commonly, agents were not always created, evolved into different states and released properly, being kept in the memory after they died. This creates problems analysing the amount and distribution of the population, which is crucial for the system in posterior optimisation steps. Some of the examples found were the following:

  - Mature agents can die and have a child in the same time step.
  - The cell was wrongly released if an agent becomes old.
  - When an agent becomes old, its corresponding mature one was not deleted properly.
  - Old agents can die by chance and die again in the same time step because they are too old.
  - Young agents can die but, if at the same time they become mature because they reach the corresponding threshold, then the system tries to create a mature agent.
  - Indirect methods to populate the city, like *Agent::MoveOrEmigrate*, did not assign a cell to the corresponding agent. Then, this behaviour creates problems when the calculation of the population size is performed and also if the agent is removed from the system after its death.
  - The implemented method, which returns the cells according to the Moore's neighbourhood, only retrieved the half of the neighbourhood.

- Errors were found in the system that generates the event-drive paradigm by including methods in the *schedule*. This capability is implemented in Repast Simphony using *annotations*. Any method that has to be called each time step automatically by the system needs to have a corresponding annotation in its declaration. Additionally, the class of the method must be included in the schedule at the starting point of the program. All methods were revisited, since some of them have annotations, but they were never called by the schedule. Concretely, methods in the classes *Demographics* and *City* included annotations but the classes were not in the schedule.

- The age of the cell was not always updated in each time step, creating inconsistencies in the evolution of the city.

#### 3.2.1.3 Implementation of extra functionality

The model was extended in multiple ways, adding further functionality in different areas. A brief summary of the most important elements are introduced below:

- The goal of the model was modified to study the construction of policies aimed at protecting ecological areas. A layer of ecological information was added to the CA. The construction of the landscape was designed to assign to each cell a stochastic value which represents its environmental value. Additionally, a mechanism of feedback was implemented to mimic the ecological influence of its surroundings over the objective area. This feedback can be positive or negative.

- The dynamic of the ecological degradation caused by urban sprawl was also included in the model. Changes in the state of the rural areas and relationships created between rural prices were also added.

- Due to the theoretical nature of the developed framework, the traditional economic model of Alonso was selected to guide the agents to select a cell to settle down. This was done by the modification of the utility function of each agent to comply with the basis of this model: a trade-off between housing prices and distance to the city centre.

- In the initial version of the model, each cell had a predefined maximum capacity related to the number of agents that could settle down on it. This capacity was fixed, equal to one, and homogeneous for all the urban cells in the lattice. This means that each cell could allocate only one agent along with their offspring. When a young agent becomes adult, it tries to find their own cell in the model. If there is no free space, the agent is removed from the model.

  This management of the population could give problems when this threshold is reached and agents have new offspring. Besides, by adding this constraint to every cell, a bias is introduced in the system, even if this factor can be represented by a random value in a defined range. Under these circumstances, the pure implementation of Alonso model could be compromised, since this restricts the capacity that cells in the city centre can reach in a short amount of time. Then, capacities where changed to a system based on growth, led by the dynamics of supply and demand of land price. The implementation of these concepts are more in line with the theory behind the economic model of Alonso, where no removal of agents due to a limited capacity of the city is performed.

- The dynamic nature of the population was increased by implementing the initial population and migration randomly.

- The initial model was a pure urban model with no management of cells located out of the bounds of the city. In the extension of the model, the characterisation of different land-use types were added. These cells were classified into forest and agricultural cells. Also, multiple price gradients that were affected by different properties of the system were included. The CA was extended from Fig. 3.2 to cover land-use types which can represent the different states of rural land. EMPTY cells are characterised by FOREST and AGRICULTURAL and a new type that represent a protected green area/park is called PROTECTED.


Figure 3.2: Basic state machine of the life-cycle of a cell in the grid without ecological characteristics, from the creation of the cell at the beginning of the simulation until the cell is urbanised.

- The size of the grid was enlarged from 25 × 25 cells to 50 × 50 cell, see Fig. 3.3. Larger configurations where rejected due to the limitations in performance of the project, which should be able to run on a conventional PC.

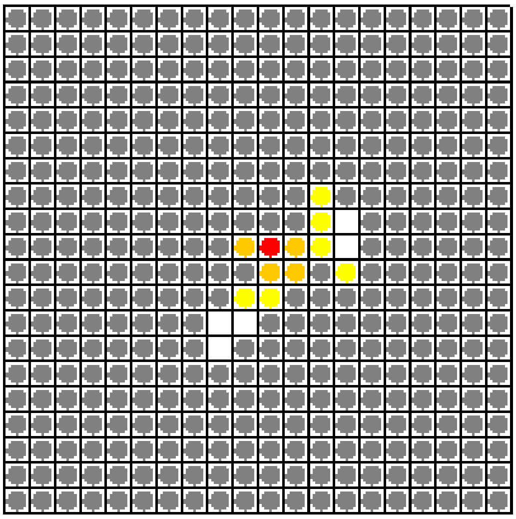


Figure 3.3: Basic state machine of the life-cycle of a cell in the grid without ecological characteristics, from the creation of the cell at the beginning of the simulation until the cell is urbanised.

## 3.3 Population Dynamics

Every time that a new family wants to settle down in the city, they should decide the best place to live according to their personal preferences and restrictions. The new families can originate from another family that was already located within the city, or by incoming migration. These incoming flows create a rising demand for housing which boosts the development of new urban areas. The idea that the city expands from its centre to balance the housing demand to its supply was proposed by Muth (1969). The economic equilibrium is achieved by the aggregated individual negotiations based only on local information.

Positive migration is attracted by major opportunities for economic growth caused by the local demand for labour. In the model, the migration factor is managed by a variable called *migrationRate* that controls the amount of adult agents that are externally included in the model. This factor is defined stochastically to allow a different and unknown number of people to arrive in each lapse of time. *migrationRate* is always positive with a maximum bounded defined by the 2.9% of the current population and an absolute limit called *maxMigration* equal to 1000.

Each cell has the capacity of allocating multiple families. There is no restriction defined in the system about the number of families living within a given cell because

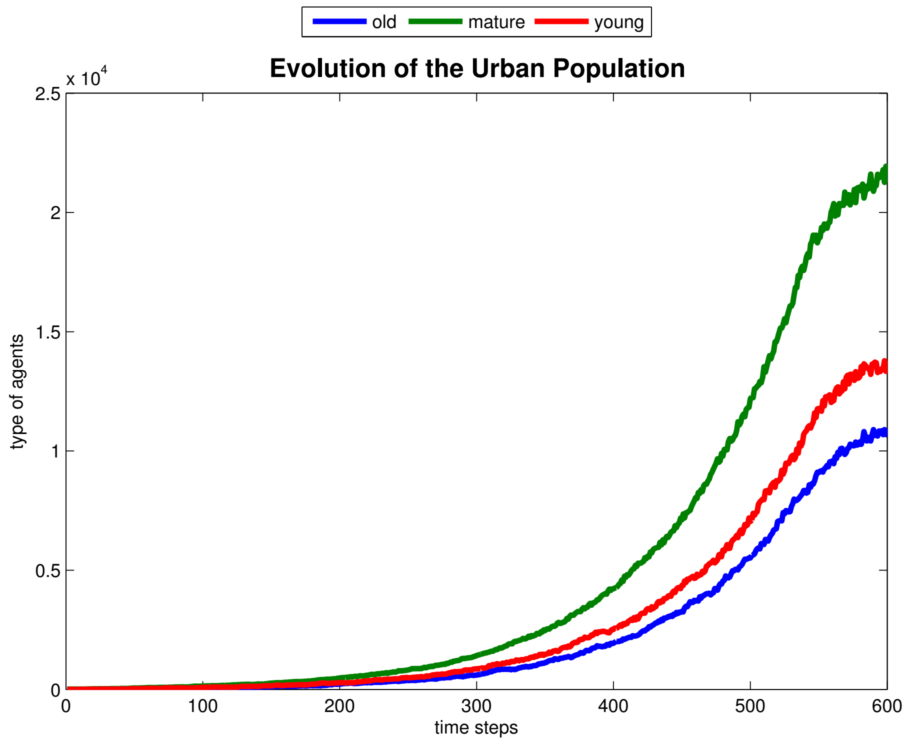



Figure 3.4: Visual representation of the population distribution using age as classification criterion. Young agents are represented in blue, adults are depicted in red and old in green.

each cell has enough capacity to allocate an unlimited number of agents. This means that the density of a neighbourhood is unbounded and it is only affected by a higher residential demand and the consequent increment in housing prices. In the long term, the high prices will drop the attractiveness of this concrete area, reducing the demand of land.

The selection of a residence involves interaction among different parameters. As in real-world cases, living in a property entails some expenses including a rental cost and transportation cost to commute daily in the model. The number of offspring is not taken into account when these costs are calculated. To afford these costs, the model is configured in such a way that each family receives a different gross household income $\alpha$. For each family this value is fixed for all the time the adult agent is employed. The function which determines the salary $\alpha$ is defined for each new adult agent $a_i$ that wants to find a new household in the city as follows:

$$\alpha(a_i) = \frac{1}{n} \times \sum_{i=1}^{n} p_t(c_i^t) \times SF \tag{3.3}$$

where every cell $c_i^t \in N_U^t$ are the elements of the set of urban cells in the system in a determined time step $t$, $p$ is the function that returns the price of an urban cell at this time $t$ and $SF$ is the constant called *Salary Factor* that is equal to 1.7. This constant value is defined based on a similar urban models developed by Wu and Plantinga (2003), where it is conceived that in the Alonso's model after commuting costs, each agent spends one-half of its budget in housing. The values chosen for their model are consistent with empirical evidence gathered from census of the USA. According to the same previous formula, since urban prices increase with time, see Fig. 3.6a, income is also positively adjusted with time.

For simplicity and following the assumptions of Alonso's model, it is assumed that there is zero unemployment for all the population at any moment of the simulation. The housing market is assumed to be in equilibrium, where all the population is allocated inside the urban area. This requires that residents optimise their choices based on the prices of other sample locations. Agents select their preferred household following the classical microeconomic equilibrium model of Alonso (1964) searching for a trade-off between their personal preferences and their economical restrictions assuming global knowledge of the current offer. Concretely, they look for the maximisation of a utility function in the pursuit of an economic competitive equilibrium between housing space and community costs in which externalities, in the form of green areas have been introduced. The introduction of externalities in the form of local public infrastructures and open spaces was previously analysed (Fujita, 1989). Thus the utility function can be formulated as:

$$\begin{gathered} \max U = (w, z, d, p : w > 0, z > 0, d \geq 0, p > 0) \\ \text{such that: } w - z - K * d + p = 0 \end{gathered} \tag{3.4}$$

where $d$ represents the distance from the household to the corresponding CBD that, in the case a monocentricity configuration is located in the centre of the lattice.

This distance can be calculated as:

$$d = \sqrt{\left((x-u)+(y-v)\right)^2} \tag{3.5}$$

$w$ is the wage received monthly calculated as shown in formula 3.3. This value does not change throughout the adult life of the agent. $z$ is the price of using the residential property, which is constant for the entire life of the agent once the value is agreed and $K$ is the constant marginal community cost with a value stored in the variable *transportRate* equal to 25. Travel cost $k$ considers only commuting from the job centre $(x, y)$ which is located in the corresponding CBD to the residential location at $(u, v)$ and it can be defined as:

$$k = K * d \tag{3.6}$$

Finally, $p$ represents the agent's preference for dwellings located close to green areas, which implies their acceptance to pay more for this kind of residential unit; this increases the demand for the nearest dwellings and, as a consequence, provokes the modification of urban patterns. Since, not all agents share the same desire to live close to these areas, $p$ can add an extra value in function of this factor and the distance to the closest green area. Concretely, a variable called *greenSpacePreference* takes a random value from 0 to *maxGreenPreference*, which is equal to 10%.

$$p = greenSpacePreference \times (4 - minGreenDistance) \tag{3.7}$$

where *minGreenDistance* return a positive value only if the distance to the green area is lower than 4.

The personal preference for green spaces $p$ is parametrised by applying a uniformly random process $\mathcal{U}(0, 1)$ over the population. Following this utility function, agents populate the urban cells of the grid. By selecting a concrete residential location households are also choosing a level of green services and amenities and a rental price.

The pseudocode of the implementation of the search that each agent performs for a household to live according to Alonso's model is the following:

**Algorithm 3.1** Agent's search for a household

```
global variables
    WAGE_FACTOR = 1.7
    PENSION_RATE = 0.9
end global variables
procedure findNewLocation
    require cell, salary
    Cell cell = null;
    salary = 0;
    if agent instanceof Mature then
        salary = ((Mature)this).getWage();
    else
        if agent instanceof Old then
            salary = salary * PENSION_RATE;
        else
            System.err.println("A child trying to find a flat on his own");
        end if
    end if                                  ▷ Select the cell according to the salary
    cell = lattice.evaluateCell(salary, agent.greenSpacePreference);
    if cell == null then
        System.err.println("Agent didn't find any available residence");
    else
        ((Mature)agent).setRent(cell.getPrice());
        cell.incrementDemand();
    end if
end procedure
procedure getWage
    require WAGE_FACTOR
    return getAvgUrbanPrice() × WAGE_FACTOR;
end procedure
```

In the algorithm 3.1, *WAGE_FACTOR* and *PENSION_RATE* are constants that represent the spent in housing and the correction factor between a salary and the corresponding pension. *cell* is the cell selected as a household for the agent and *salary* is the salary of the agent. The complexity of the function is $\mathcal{O}(n)$, where $n$ is the number of individuals that are searching for a new household in each time step.

### 3.3.1 Agent's Life cycle

During its entire life in the system, each agent has associated a unique state value that can change through time. The possible set of state values defined for the population includes: YOUNG, ADULT and OLD and the complete flow that allows the transitions between them is depicted in Fig. 3.5.

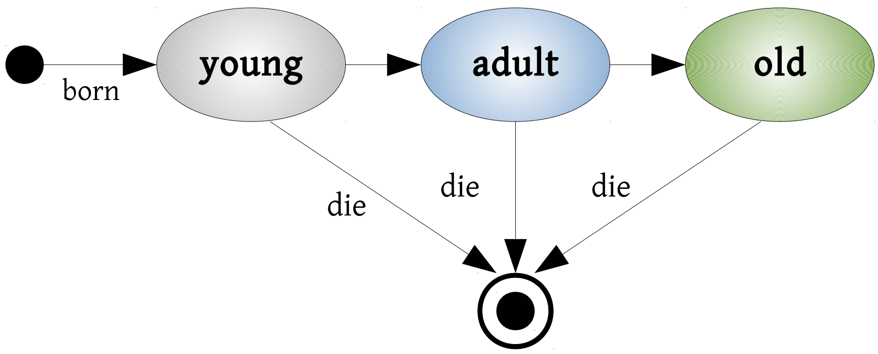



Figure 3.5: Life-cycle of an agent

The evolution of the states of each entity depends exclusively on the age of the agent and on the possibility that the agent dies this turn. The agent can die in each state, but the probability is higher when the agent is older. The concrete values are 0.1% for young agents, 2% for adults and 8% in the case of old agents. If the agent dies, all their children and themselves will disappear from the model. An agent is not capable of living more than 100 years.

Agents are capable of having offspring when they become adult; that occurs when young agents reach 20 years old. At this time they leave their parents searching for their own dwelling. The probability of having a child, which is defined by the parameter *BirthRate* is 4.5% in each year. There is no limitation on the number of possible children that each agent can have, but they only have the potential of being a parent during adult life. Once the agent achieves 50 years old and becomes OLD this capability disappears. The idea behind the transition to an old status is that this agent retires. At this point the agent does not have to pay transport costs any more and additionally its salary is reduced by a 10%.

In Figure 3.4 the distribution of the population at a determined instant of the simulation is plotted. There we can distinguish the evolution through time of the total number of agents settled down in the city according to their age, grouped into three kinds of citizens: young, adult and old.

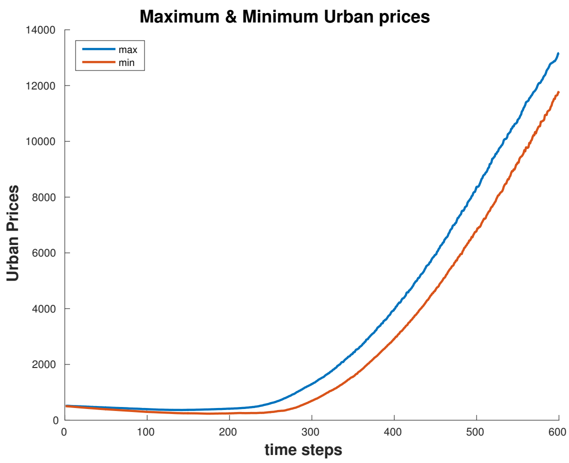
Table 3.1: Population Variables

| **General Population Dynamics** | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| migrationRate | 2.9% of current population | External incoming population coming from migration |
| maxMigration | 1000 | Maximum value of migrationRate |
| salaryFactor | 1.7 | Amount of salary invested in housing each month |
| **Agent Dynamics** | | |
| **Name** | **Value** | **Description** |
| YOUNG | $<20$ | Agents are young when their age $< 20$ |
| ADULT | (20-49) | Agents are adult meanwhile their age is between 20 and 49 |
| OLD | $> 49$ | Agents become old when age $> 49$ |
| death(young) | 0.1% | Likelihood of death for young agents |
| death(adult) | 0.2% | Likelihood of death for adult agents |
| death(old$<$100) | 0.8% | Likelihood of death for old agents with age $< 100$ |
| death(old=100) | 100% | Likelihood of death for old agents with age $= 100$ |
| offspring(age$<$20) | 0 | Likelihood of having offspring for agents with age $< 20$ |
| birthRate | 4.5% | Likelihood of having offspring for agents with age between 20 and 50 |
| offspring(age$\geq$50) | 0 | Likelihood of having offspring for agents with age $\geq 50$ |
| transportCost(old) | 0 | Retired agents do not have to pay transport costs since they do not require to work |
| salaryReduction | 10% | Retired agents suffer a reduction in their income |
| greenSpacePreference | ran(0-9) | Quantification of the preference of an agent to live close to a green area |
| maxGreenPreference | 9 | Maximum boundary for the variable greenSpacePreference |

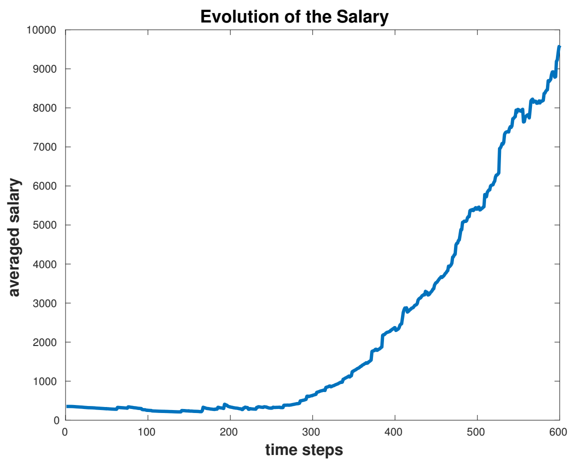
A summary of the variables used to model population dynamics are described in Table 3.1. Values described in this section were selected based on empirical analysis. Test and error techniques where used to mimic the desired behaviour of the model.

## 3.4 Urban Dynamics

Housing is a complex, multidimensional and abstract commodity which comprises multiple elements. In the present model, houses are represented by urban cells. Urban cells are a type of landscape that has been transformed from native ecosystems into impermeable surfaces. This transformation occurs when cells receive the permission to be urbanised, which figuratively means that dwellings are constructed. Once dwellings are built, they can allocate population who evolves at an aggregate level. In the real-world, there exist specific factors that trigger the construction of new developments and consequently cause urban growth. Among them, the phenomenon by which a noticeable number of people prefer to reside in one specific area of a city is selected as the most important driver in the model.




(a) Maximum and minimum bounds of the urban prices depicted through time.




(b) Evolution of the average salary of the population during the time horizon of the simulation.

Figure 3.6: Urban prices and salary share the same growth tendency to allow the settlement of new families in the model.

In each time period, different flows of incoming population search for a suitable household to settle down, causing an increment in the demand for the development of new urban areas. These urban needs can be fulfilled by the transformation of new patches of land which exert pressure over the remaining open areas located in the fringe of the city.

Once an area is selected for urbanisation and the infrastructure is created, it remains in this state until the end of the simulation. Even if redevelopment is included as a possible change of state within the urban cells, the model does not allow that the current urban land areas can lead back to open space. The reason behind this is that, in practice, it is very likely that once an area is urbanized it typically remains in that way (Nalle et al., 2002; Silberstein and Maser, 2013).

### 3.4.1 Urban Cell Life Cycle

In Fig. 3.7, each circle represents the different states allowed for each patch of land represented in the model. The defined states include: EMPTY (FOREST or AGRICULTURAL), AVAILABLE, NEW, OLD and PROTECTED. The directed arcs or connections correspond to possible valid transitions among them, which summarise how the state of the cells can change over time. All transition rules are probability-based except PROTECTED, in which its selection is based on the application of a range of more advanced methods. The selection of the final procedure depends on the current optimisation baseline applied within the model.

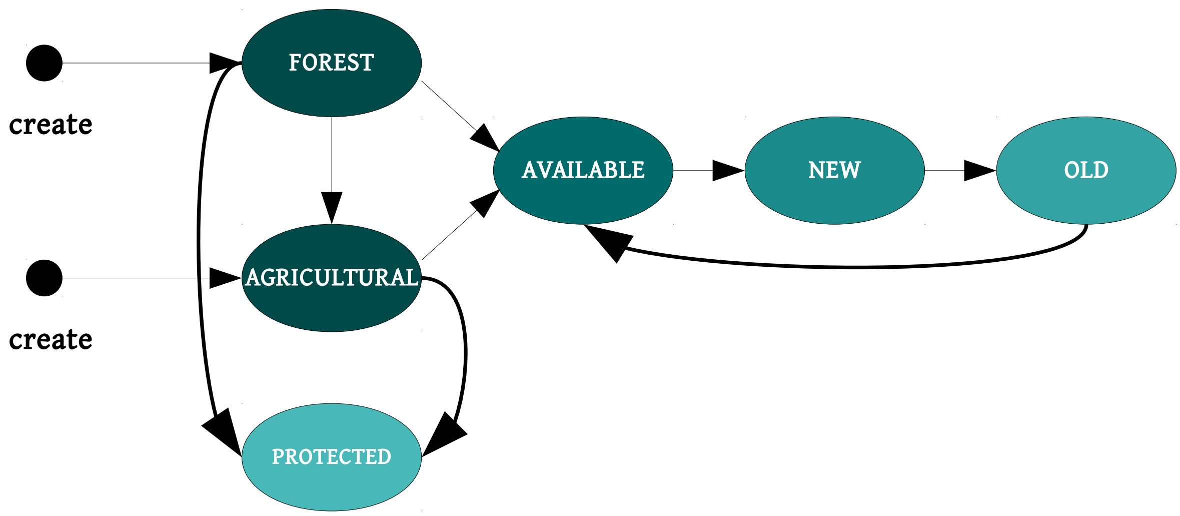



Figure 3.7: State machine of the life-cycle of a cell in the grid, from the creation of the cell at the beginning of the simulation until the cell is protected or urbanised.

All cells start the simulation with an EMPTY state that represents the concept that the area has not been urbanised and maintain more nature-like characteristics. This state can be subclassified into AGRICULTURAL or FOREST according to its environmental value. This value is assigned globally at the beginning of the simulation but can vary with the time. The evolution of this values are related to a degradation process that it is provoked by urban sprawl.

When an empty area is close enough to an urbanized cell, at distance equal to one, it can be transformed into the state of AVAILABLE with a probability of 10%. This change of state represents transitions between open areas to terrains where urbanisation is permitted but its development was not carried out yet. Once a patch of land is available for urbanisation, any other type of activity is not allowed under this land any more.

The process to construct new urban settlements, meaning the change of the status of the cell form AVAILABLE to NEW, is restricted by the constraint that the group of urban cells that comprise its neighbourhood must have at least one family living on it. This requirement is imposed to control the exponential growth of the city where, by means of the current transition rules, new non-populated areas were constructed without being required from the current urban population needs. Once this condition is met, the cell status can be changed with a probability of 30%.

Once that the cell is finally populated by a family who selected it as its best choice to live, the cell changes finally to an OLD state. Identically, when an OLD cell achieves an age threshold it can be renewed and become again NEW. This stochastic process starts with an age of the cell of 30 and it is checked and applied annually with a probability of 20%. In the renewal process, all the agents allocated on the cell are evicted and they should look for another placement to live.

### 3.4.2 Growth Behaviour of the City

In Fig. 3.7 it is illustrated how the urban cells defined in the model, denoted in the equation 3.2 by $N_U$ can be sub-divided into three categories namely available, new and old.

$$N_U^t = N_A^t + N_N^t + N_O^t \tag{3.8}$$

where $N_A$ ($available$) represents patches of land available to be urbanised, $N_N$ ($new$) depicts cells that have been just built and $N_O$ ($old$) are units areas within the city with developments constructed and inhabited in the past. Consequently, equation 3.2 can be rewritten by adding more specificity to the $N_U^t$ component as follows:

$$N_A^t + N_N^t + N_O^t + N_R^t + N_P^t = NC \quad \forall t = 0, 1, 2, \ldots, T \tag{3.9}$$

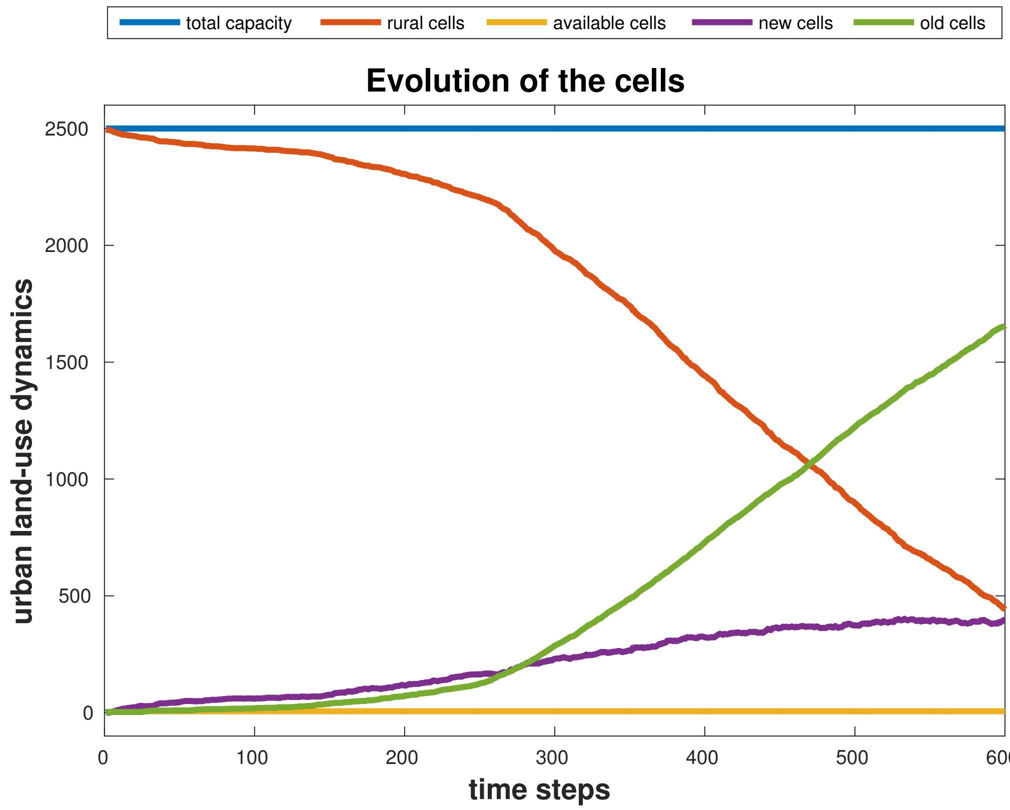



Figure 3.8: Evolution of the state of the urban cells within the lattice for the entire length of the simulation. These dynamics correspond to the system configured under a normal level of demand.

In this formula is illustrated an important characteristic of the growth dynamic, which is its one-directional nature. Available patches of land transform to new development and, from this, to established development in such a way that $N_R(t) \rightarrow N_A(t) \rightarrow N_N(t) \rightarrow N_O(t)$. The diffusion effect provides available land to the new development by using the adjacent available Moore cells of its neighbourhood. The other transition leads to a mature state when new development is surrounded by new or old developed land.

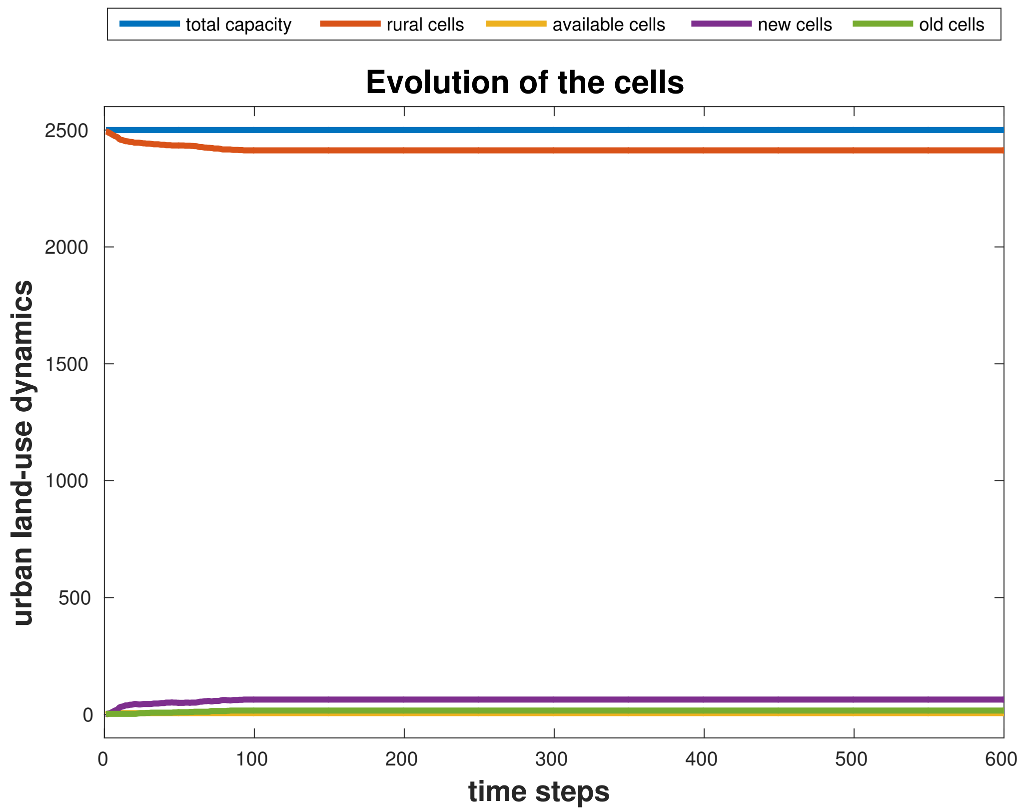



Figure 3.9: Evolution of the state of the urban cells within a lattice that represents a city without enough demand of new urban development. The number of new cells transformed is not enough to make the city grows, which implies the fall into a stasis state.

New developments are the consequence of an expected economic profit resulting from the future demand of households who desire to move to these premises. These households should find these new settlements more interesting than the available inner spaces previously built in terms of economic, number of services offered, social conditions or amenities. In this regard the developed model includes three different methods to populate the city with new adult agents:

- The initial population of adult agents. The number of individuals varies in each simulation. The range of values assigned goes from 1 to *initialMatures*, that is equal to 5.
- A variable amount of emigration of adult people coming to the city in each time step, the concrete number of which depends on the current population as

specified in the following formula:

$$M_t = \Big(\sum_{i=1}^{N^t} a_i\Big) \times migrationRate \tag{3.10}$$

where $M_t$ is the number of new agents coming to the city to settle down in time step $t$. The total number $N_t$ of urban citizens of the city $a_i$ at time $t$ is used as a metric to calculate the migration rate, due to the fact that the model assumed the hypothesis that as the population within a city grows, more job opportunities are expected to appear and this is one of the positive factors to consider in migration (Todaro, 1969). Finally, *migrationRate* is a constant that links the population with the amount of migrants and it is equal to 0.029.

- The new offspring of the adult population of the city. This value is stochastic and its dynamics depends on the total amount of adults.

In general, population dynamics in the system can be summarised in the following formula:

$$P_t = P_{t-1} + NB_t - D_t + M_t \tag{3.11}$$

where current population $P$ at time $t$ depends on the population at time $t-1$ plus the newborns $NB$, the input flow due to migration $M$ and the elimination of agents that have died from the system $D$.

If the city does not receive enough new citizens by migration mechanisms or by inner new births, the growth of the city can be severely limited, not requiring new urban developments due to a lack of demand. In this case, the scarce urban growth behaviour could be depicted by Fig 3.9.

The pseudocode of the implementation of the transition rules used in the urban model is depicted in Algorithm 3.2.

In the variable section, the constants used to transition between urban states are depicted (*RELEASE_PROBABILITY*, *DEVELOPMENT_PROBABILITY*, *REDEVELOPMENT_RATE* and *REDEVELOPMENT_AGE*). Additionally, *state* and *nextState* represent the current and the future state of the cell for the next time step of the simulation.

**Algorithm 3.2** Implementation of the transition rules in the urban model

```
global variables
    RELEASE_PROBABILITY = 0.1
    DEVELOPMENT_PROBABILITY = 0.3
    REDEVELOPMENT_RATE = 0.2
    REDEVELOPMENT_AGE = 30
end global variables
procedure UPDATE
    require state, nextState
    switch cell.state do
        case 'EMPTY'
            if nextState ! = CellState.PROTECTED then
                if getDevelopedNeighbours() > 0 &&
                Rand.nextDouble() < RELEASE_PROBABILITY then
                    setState( CellState.AVAILABLE )
                    urbanisedCell(cell)
                end if
            end if
        case 'AVAILABLE'
            if DEVELOPMENT_PROBABILITY < Rand.nextDouble() &&
            !Lattice.isCellsEmpty() then
                setState(CellState.NEW );
                build();
            end if
        case 'NEW'
            if this.agents.size() > 0 then
                setState(CellState.OLD);
            end if
        case 'OLD'
            if age >= REDEVELOPMENT_AGE) then
                if Rand.nextDouble() < REDEVELOPMENT_RATE then
                    redevelop();
                end if
            end if
end procedure
```

### 3.4.3 Prices of Urban Cells

According to Alonso's conception, urban land prices increase as long as they get closer to the CBD where jobs are concentrated and the population prefers to live in places that avoid commuting costs. The price of an urban cell represents the amount of money that agents have to pay regularly as a rental cost or mortgage, which can represent half of their income after commuting costs on average (Wu and Plantinga, 2003). In the model, the price of each household varies with time and is dependent on other factors that were included as an extension of the model. The following list includes the elements considered:

- **The Demand:** The demand for certain preferred locations causes the increase of their price. The demand is defined according to the number of families living in a given cell. According to this general conception, the dynamic behaviour of the urban prices can be summarised in the following points:

  At the beginning of the simulation the city is comprised of a single cell with the status of *NEW* which is empty of agents. The initial rent for the first habitable settlement in the system is a constant value called *initUrbanPrice* always equal to 500 monetary units. These areas are the initial CBDs in the system and its number depends on the configuration of the system.

  When the city increases its population and a new agent wants to find a new household to live in, this action can occur under different circumstances. If the agent is the first to be allocated in the cell, the price of the cell is calculated taking the average of the urban cells, which form its neighbourhood. Since the city grows by diffusion mechanisms there is always at least one urban cell in its neighbourhood.

  Let $\Omega_{c_i}$ be the function which represents the neighbourhood of a determined cell $c_i$, that is the square of $3 \times 3$ cells that compounds a Moore neighbourhood of range $r = 1$. This configuration comprises the $R = (2r + 1)^2 - 1$ cells surrounding the objective cell $c_{(x_0,y_0)}$ in such a way that:

$$\Omega_{c_{(x_0,y_0)}} = \{(x, y) : |x - x_0| \leqslant r, |y - y_0| \leqslant r\} \tag{3.12}$$

  Then, the price $p$ for an agent who selects an empty urban cell $c_i \in N_N$ to live in is the following:

$$p(c_i) = \frac{1}{R} * p\left(\Omega_{c_i}\right) \quad \forall c \in N_O \tag{3.13}$$

where $R$ is the constant which represents the number of cells of the neighbourhood and $N_O$ is the set of urban cells with developments constructed. After the price of the cell is calculated, the updating procedure does not use again the information about the neighbourhood of the cell, instead it is based on the posterior demand of the location which follows these rules:

- For each new family who wants to live in the same area $c_i$, the current price is increased due to an increment in the demand for this particular location.

$$p(c_i)_t = p(c_i)_{t-1} \times urbanDemandPriceStep \tag{3.14}$$

  where *urbanDemandPriceStep* is equal to 0.5%.

- In turns where a family moves to another location or dies and its dwelling is released, the demand decreases slightly and consequently the new price for this area is also reduced:

$$p(c_i)_t = p(c_i)_{t-1} \times lessPopulationPriceStep \tag{3.15}$$

  where *lessPopulationPriceStep* is equal to 0.1%.

- Finally, if there is a continuous drop in the demand on this cell for a period of time, which is defined in the model when the cell is undemanded for more than three time steps (*maxTimesUndemanded*), then the price is further decreased by:

$$p(c_i)_t = p(c_i)_{t-1} \times undemandedPriceStep \tag{3.16}$$

  where *undemandedPriceStep* is equal to 1%.

- **The proximity to a green area:** The proximity of a green area is a factor which affects the final price of the houses. A green area is considered close enough to modify the prices of the houses if the distance to any green area

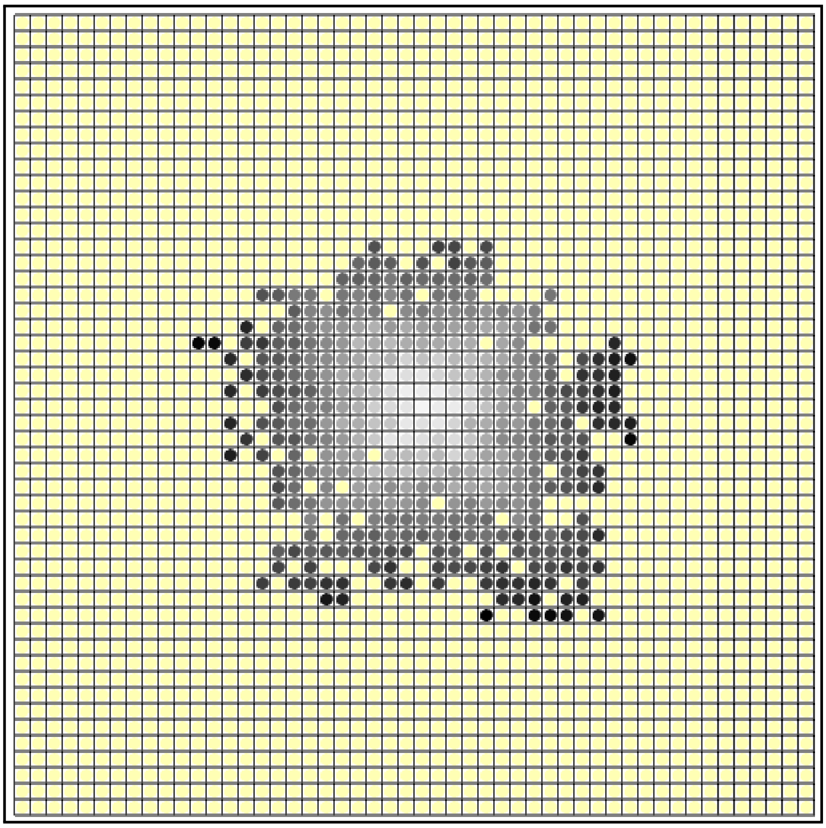


Figure 3.10: Distribution of urban prices in a monocentric configuration of the model. White values represent cells with the highest price and black areas depict the cheapest dwellings in the model for time step equal to 300.

is between 1 and 3 cells. A visual representation of this area of influence is depicted in Fig. 5.4. Currently, this factor increases the money each agent is willing to pay for their dwelling by a maximum of 10%. See Eq. 3.7 for more details.

Once the price of the cell is agreed for a given agent, the model assumes that landowners cannot change the prices of its property while the same family is living in this dwelling.

Finally, as a consequence of the application of these rules, urban housing prices are determined in a spatial market equilibrium. Fig. 3.10 shows the final spatial pattern created by this urban prices configuration in a city composed by a single CBD.

Table 3.2: Land Variables

| Name | Value | Description |
|---|---|---|
| **Urban Cells** | | |
| **Name** | **Value** | **Description** |
| AVAILABLE | 10% | Likelihood an EMPTY cell changes to AVAILABLE |
| NEW | 30% | Likelihood an AVAILABLE cell changes to NEW |
| OLD | #agents $\neq 0$ | A NEW cell changes to WHEN when any agent settles down on it |
| redevelopmentRate | 20% | Likelihood of redevelopment of the cell if age of the cell > 30 |
| **Price of Urban Cells** | | |
| **Name** | **Value** | **Description** |
| K | 25 | Rate of change in prices of land in function of the distance from the CBD |
| initUrbanPrice | 500 | Initial urban price of the cells selected as CBDs |
| demand | 10% | increment of the price due to the demand |
| demandPriceStep | 0.5% | Increment the price of the cell for each family who select it to live |
| lessPopulationPriceStep | 0.1% | Drop in the price when a family leave a cell |
| undemandedPriceStep | 1% | Drop in the price of a cell when it is undemanded for a period of time |
| maxTimesUndemanded | 3 | Times a cell needs to be undemanded to trigger a drop in its price |
| **Rural Cells** | | |
| **Name** | **Value** | **Description** |
| forestLimit | 0.7 | Minimum ecological value of a forest |
| urbanLimit | 0.3 | Maximum ecological value of an urban area |
| bioValueStep | 0.01 | Feedback value added to a cell due to the ecological influence a neighbour |
| **Price of Rural Cells** | | |
| **Name** | **Value** | **Description** |
| $P_e$ | 100 | Profitability estimator of the earning received from the transformation and selling of urban land |
| $P_{bc}$ | 5 | Corrective term of the weight that forest and agricultural land have in the final price or a rural cell |

A summary of the variables used to model population dynamics is described in Table 3.2. Values described in this section were selected based on empirical analysis. Test and error techniques where used to mimic the desired behaviour of the model.

## 3.5 Rural & Ecological Dynamics

Non-urban cells are the type of cells that have not undergone any urbanisation transformation. These rural areas can be defined by their ecological characteristics and their current price. Both concepts, ecological value and price, are also inter-related in such a way that the price could change if the current ecological characteristics are externally modified. According to the ecological value, these cells are divided into forest and agricultural cells. The agricultural areas are patches of land with arable land and forest cells are represented by stands of trees with higher ecological value.

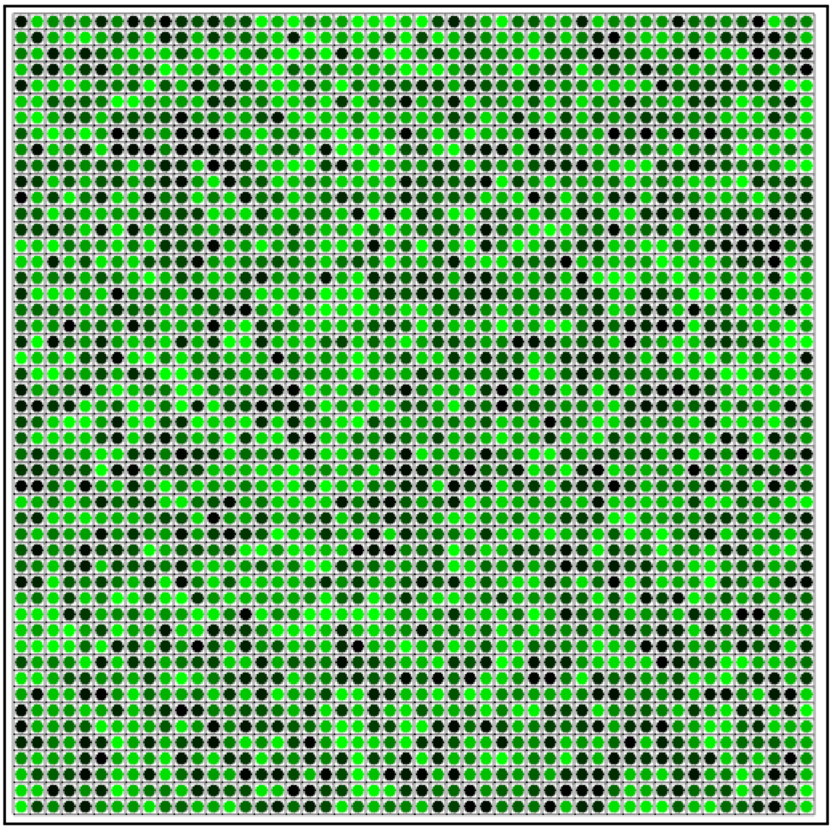


Figure 3.11: Initial ecological configuration of random environmental values assigned to each parcel of land in the grid. Light green cells represent areas of the lattice with the highest ecological values, meanwhile patches of land represented by black colour depicts areas of low ecological interest.

### 3.5.1 Ecological Value of the Cells

At the beginning of the simulation, the model assigns a stochastic value called *ecoValue* to the entire set of non-urbanised cells. This parameter, generated by a uniformly random process $\mathcal{U}(0,1)$, represents the ecological value of this parcel of land. A visual representation of the generation of these values in the grid can be seen in Fig. 3.11.

**Algorithm 3.3** Feedback dynamics of the ecological values within the model

```
global variables
    FOREST_LIMIT = 0.7
    URBAN_LIMIT = 0.3
    BIOVALUE_STEP = 0.01
end global variables
procedure ECOVALUE
    for each cell c in lattice do
        ecoValue(c) = ran(0,1);
        for each cell c_n in neighbourhood(c) do
            if ecoValue(c_n) >= FOREST_LIMIT then
                ecoValue(c) += BIOVALUE_STEP;
            end if
            if ecoValue(c_n) <= URBAN_LIMIT then
                ecoValue(c) -= BIOVALUE_STEP;
            end if
        end for
    end for
end procedure
```

In the algorithm 3.3, *FOREST_LIMIT* and *URBAN_LIMIT* are constants that represent the minimum ecological value of a forest land and the maximum value of an urban area respectively. *BIOVALUE_STEP* is a value added as a positive or negative feedback to a cell due to the influence of the cells that belong to its neighbourhood. All of these values where manually selected based on the empirical testing of their effect on the model. The complexity of the algorithm is $\mathcal{O}(n^m)$, where $n$ is the number of cells in the lattice and $m$ is the number of cells in their neighbourhood.

However, the final measurement of the ecological richness of a patch of land is not an isolated concept. This value is also influenced by the ecological values of the surrounding land. In this case, this concept is represented by the cells that form its neighbourhood and the positive or negative influence of each cell is represented by the constant *BIOVALUE_STEP*, which is initialised to 0.01. The implementation of the generation of these values is shown in the algorithm 3.3.

The final ecological value of a cell stored in the *ecoValue* variable is used to identify different rural land-use types within the model. If the concrete value is higher than the value of the constant *FOREST_LIMIT*, which is 0.7, then this cell is classified as a *forest* cell, otherwise it is considered *agricultural*. The belonging to each category is dynamic over the time.

The ecological dynamics of the model are mainly characterised by an uninterrupted bio-degradation process due to the continuous urban expansion and the transformation pressure on the peri-urban areas of the city. These urban transformations provoke changes in the rural land use due to the spread of the negative feedback of these urban areas to the surrounding, which cannot have an ecological value higher than 0.3 (threshold value stored in the constant *URBAN_LIMIT*). Then, if any preventive protection measures are implemented, rural cells tend to transform with the time from *forest* to *agricultural* state.

### 3.5.2 Prices of Non-Urban Cells

Transformation of rural patches of land to be urbanised normally involves important capital gains. The first studies which tried to analyse this process used models constructed on the basis of Ricardo (1891) capitalisation formula, $P = R/i$, where land price is calculated with the use of the discounted differential rent. $R$, in this case, is the annual differential rent or the absolute rent and $i$ is the discount rate (Arnott and Lewis, 1979). However, due to its simplicity the validity of the method has been questioned (Weersink et al., 1999). Other methods are based on hedonic price equations (Chicoine, 1981), land rent gradients (Capozza and Helsley, 1989; Plantinga and Miller, 2001) or gravity models (Shi et al., 1997).

Furthermore, even if land revenue in the calculation of rural land prices is a significant factor, it is not the only determinant to take into account (Latruffe and Le Mouël, 2009). Ricardo's approach is based on the assumption of the existence of a competitive market. The reality, however, is that externalities and market imperfections are important factors to influence the final land price (Ciaian et al., 2012). This is the case of non-farm investors (Shiha and Chavas, 1995), who increment cost transactions and the regulative role of institutions, which create agricultural subsidies, environmental regulations and other land regulations (Latruffe and Minviel, 2012). The problem is that little is known about how these types of regulations of land transactions affect rural land prices (Latruffe et al., 2013).

These economical models assume population density growths and the extension of urban boundaries. In peri-urban areas non-urban prices fall with distance from cities due to the expected substantial capital gains from a future urban development. These land rents are determined not only by intrinsic agricultural factors but also by the potential future residential rent and the time horizon where the transformation from agricultural to residential use is planned.

Non-urban prices have opposite behaviour in comparison with the rural counterpart. Rural land dynamics are more expensive when they are located in the physical boundaries of the city, decaying with distance. Plantinga and Miller (2001) postulated that agricultural prices are influenced by the agricultural exploitation and by the expected future urban transformation profitability. For our purposes and based on a Plantinga's simplification of formula 9, the final price of a protected cell, $P_t^P(z)$ located in cell $z$ with coordinates $(x, y)$ at time $t$ is the following:

$$\begin{aligned} \rho_t^U(z^*(t)) &= P_t^U(l) \cdot P_e \\ P_t^P(z) &= \Big(\frac{P_{base}}{P_{bc}} + \rho_t^U(z^*(t))\Big) \cdot e^{-\alpha[z-z^*(t)]} \end{aligned} \tag{3.17}$$

where $P_t^U(l)$ is the price of the urban cell $l$ most recently urbanised in time $t$, $\alpha$ is the *Change Rate* that measures the declining urban rent gradient from the CBD. $P_e$ is the profitability estimator that measures the earnings from transforming and selling a rural cell into an urban area. $z - z^*(t)$ defines the physical distance from the cell $z$ to $z^*(t)$ where $z^*(t)$ is the placement of the peri-urban area at time $t$ such that

Table 3.3: Prices per cell

| Type of Cell | Area x Price | Final Price |
|---|---|---|
| 1 cell forest | 51.8 ha x £6,600 | = £854,700 |
| 1 cell agriculture | 51.8 ha x £3,000 | = £388,500 |

$z > z^*(t)$. $\rho_t^U(z^*(t))$ depicts the estimation of the profitability of the future urban transformation. Finally, $P_{base}$ is based on rural land prices (agricultural and forest) in the UK (Riley, 2002), see Table 3.3 and $P_{bc}$ is a corrective factor that controls the importance of the $P_{base}$ term in the final price of the land.

Apart from the price dynamics at a given time, prices of the entire set of non-urban cells increase with time as the total number of rural cells becomes less numerous due to urbanisation and the consequent decrease in supply.

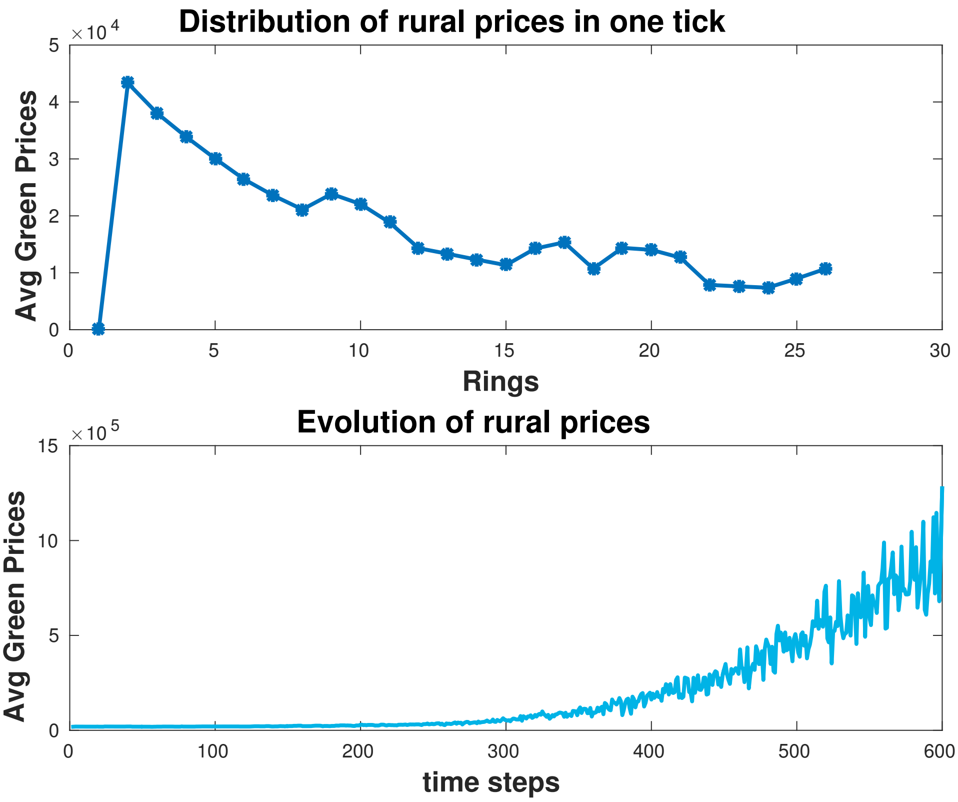



Figure 3.12: Rural Prices dynamics: Distribution of the prices within the lattice in a single tick of the clock. The grid is divided into concentric annuli or rings and rural prices are averaged accordingly (top). Rural prices of the entire lattice are averaged for each tick of the clock (bottom).

In Figure 3.12 is depicted the evolution of non-urban prices during the simulation period. The critical growth in prices that occurs at the end of the simulation is noticeable. This is due to the the demand for a scarce number of available cells drastically raises their value.

Some researchers include other variables like the parcel size. Cavailhès and Wavresky (2003) calculates that the price-size relationship can be determined by the following characteristics: prices of large parcels, more than 1 Hectare, are constant around €3000. Prices slopes steeply increase when the size is close to zero.

Agricultural economics states that large parcels of land entail higher productivity. However, in real market conditions, prices of agricultural land per hectare decline with size. To explain this dichotomy Colwell and Munneke (1997, 1999) investigated the relationship between urban land prices and sizes and they concluded that the structure of land prices is convex as a result of the subdivision costs (Colwell and Sirmans, 1978).

### 3.5.3 Degradation Process

The formula used to reproduce the ecological degradation effect due to urban expansion is applied over all the non-urban cells at the same time as the transition rules. The equation used is summarised as follows, $\forall c \in N_R$:

$$EcoValue(c)_{t+1} = EcoValue(c)_t - \begin{cases} 0.1 & \text{if } \delta(a, \bar{x}) = 1 \\ 0.05 & \text{if } \delta(a, \bar{x}) = 2 \\ 0.01 & \text{if } \delta(a, \bar{x}) = 3 \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

Where $\delta$ is the distance from the cell $c$ to the boundaries of the city $\bar{x}$. This dynamic can change the state and the *priceBase* of the non-urban cells that are closely located to the city and hence, can influence the purchasing process of protected areas that it is restricted to our current *budget* (see Table 3.3).

Once the entire set of *BioCellValue* parameters are updated, the algorithm 3.3 is run again to update the ecological feedback processes among cells. In this way, cells located close to the boundaries of the city are more affected by the degradation

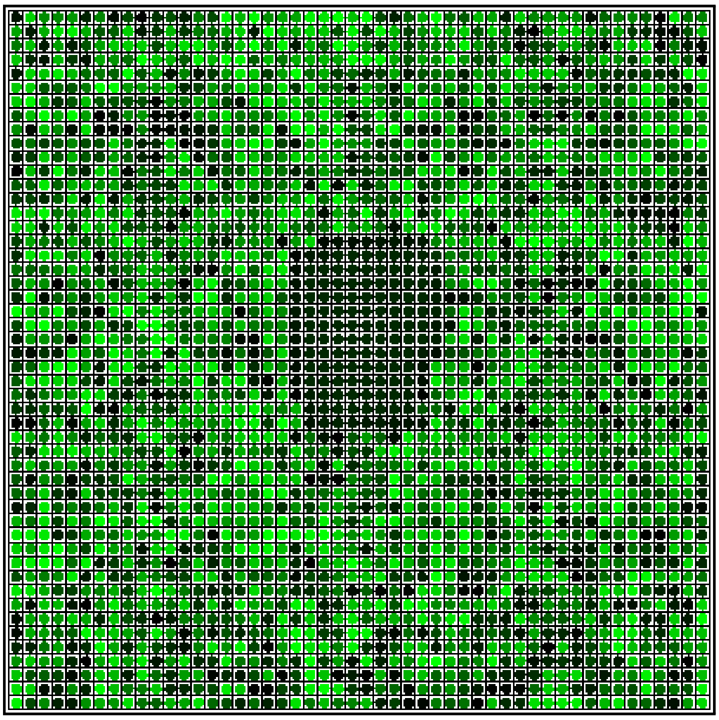


Figure 3.13: Environmental values and the effect of the urbanisation process in the grid. The range of colours from green to black depicts the ecological values of the cell. Notice that in the centre where the city is located, the black eco-values represent the biological degradation of the metropolitan area.

process in the cells which form the peri-urban belly of the city.

### 3.5.4 Protection for Conservation

In a free market scenario, open spaces can be classified as a kind of public good. This type of facility has always been associated with the cases of non-optimal markets characterised by the appearance of several market failures(Weimer and Vining, 2005), such as externalities and information asymmetry, which implies inefficient resource allocation and under-provision in the absence of any kind of government intervention(Tolley, 1974; Alterman, 1999; Kotchen and Powers, 2006). Available space is generally limited in city centres, and the application of a set of concrete policies can mitigate the continuous pressure to convert the available open space into building sites (Glickman, 1999) which is a typical consequence of urban sprawl.

Public open space planning allows local authorities to boost the protection of certain areas from the urbanization process, and foster welfare and the formation of healthier urban environments (Gillham, 2002). Real-world examples of

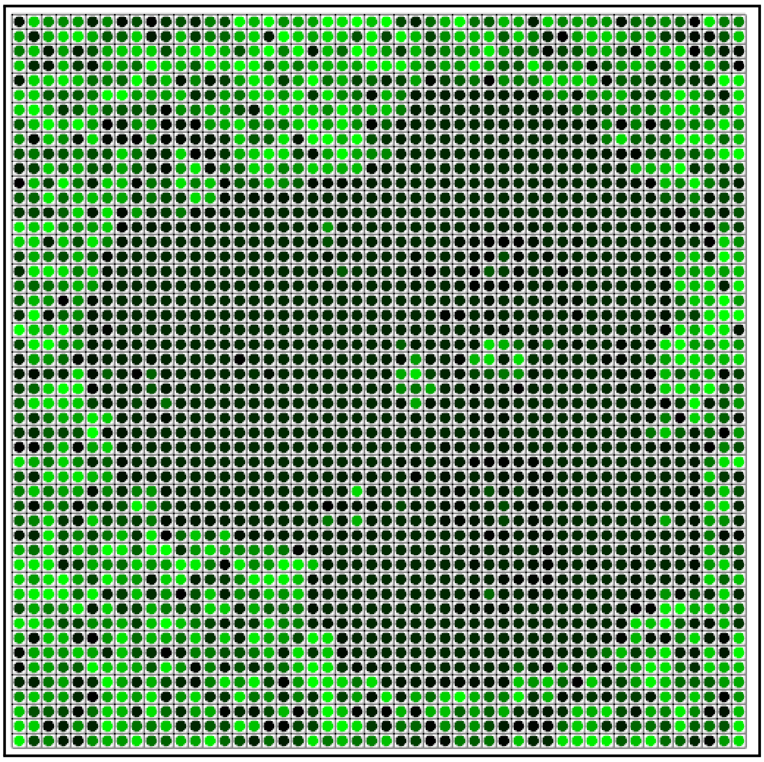


Figure 3.14: Protection effect caused by the location of open areas within the city. The figure shows the ecological values of an urban area, illustrating the role of protectionism which is able to maintain some valuable ecological areas in the centre of the city.

planning strategies based on public-ownership are Australia, Sweden, China and India (Acharya, 1987). Based on the premise that public institutions can play an important role in the protection of green areas by means of acquiring parcels of land to be transformed into parks, the model delegates the responsibility of selecting the best non-urbanised stands to a new special agent called *Municipality*. This special agent does not interact with the rest of the agents as usual. Instead, their main goal consists of managing the open space protection policy to support land conservation by the sequential purchase of green areas within the city, using a monetary income called *budget* received periodically and should be used effectively.

The green space provision is performed by a land banking mechanism where the purchase of parcels is done in advance, according to the expected population distribution and density projections. This approach was also included in other models that investigate the allocation of green areas (Wu and Plantinga, 2003). From a land banking perspective, and in order to provide the services of the foreseen residents that will be located in old or new neighbourhoods, land purchase decisions could be mainly based on the assumption that the economic benefit achieved buying non urban land far from the CBD can lead to a qualitative and quantitative improvement in the future green configuration of the city once the city grows.

The planning strategy can follow multiple approaches. According to Maruani and Amit-Cohen (2007), public open space planning can be divided into two main categories: a demand and a supply approach. A demand approach assumes that people get benefit from the presence of recreational areas in the surroundings of their residences. In a supply approach the strategy aims at protecting the existing landscape and biodiversity from an ecological point of view. The selection of the areas should be performed prioritising high-quality and uniqueness environmental properties in an attempt to preserve the most vulnerable areas. Prices of parcels, ecological value of the land, distance to the boundaries of the city along with their availability are the major factors to consider when the land purchase strategy is planned.

This selection process can be formulated as follows: $C$ is defined as the finite set of cells included into the lattice, $R$ the subset of rural cells both agricultural and forest, each of them with a corresponding purchasing cost $\alpha$ that varies with the time. $P$ the subset of cells that are protected and $U$ the urbanised cells such that $\{R, P, U\} \subset C$ and $R \cap P \cap U = \emptyset$, then the condition that a candidate cell has to hold in a given time step $t$ to be included into $A$ that is the set that contains the candidate cells available for purchasing such that $A \subset R$, can be defined as:

$$\begin{gathered} \forall \text{ cell } c \in C \\ \text{if } \alpha(c)_t < \text{budget}_t \wedge c_t \notin \{P, U\} \implies c_t \in A \end{gathered} \tag{3.19}$$

Once the candidate set is defined, the purchasing and protection phase can be formalised as:

$$
\begin{aligned}
&\forall \text{ cell } a \in A,\ \underset{strategy}{max}\ \delta(a)_t \\
&\implies a_t \in P \wedge a_t \notin A \wedge update(budget_t)
\end{aligned}
\tag{3.20}
$$

The function $\delta$ represents the metric that measures the level of accomplishment of the policy. Every subset of selected cells has associated a level of success. The model should select the configuration of green areas which achieves the highest possible level of reward according to the restrictions of the system during the considered period of time.

Once the purchase is concluded, the state of the cell is changed to *protected* and the future construction of urban facilities within its boundaries is forbidden. An example of the effect of ecological protection in the model is shown in Fig. 3.14.

## 3.6 Conclusions

In this chapter, it has been presented a multi-featured computational model of urban growth, which builds on Alonso's theoretical framework for urban growth and land economics. The developed computational model comprises a Cellular Automaton and an Agent-based System, which work together to simulate the dynamics of a city that grows in terms of both, population and land development.

The model incorporates and reproduces a varied range of interacting processes. One such process, forming the focus for this thesis, is the nexus of activities which lead to, and follow on from, the acquisition and protection of a new green space. This aspect of the land-use forms the key point of interaction between optimisation algorithms and the model. The further application of this model for planning tasks and optimisation is the core topic of later chapters. Concretely, in the next chapter, it is considered how optimisation methods, when tasked with discovering what inputs to the model will lead to the best outcomes, can be engineered to be able to handle the non-linearities and uncertainty inherent in such a complex model.

# Chapter 4

# Evolutionary Algorithms under Noise and Uncertainty

## 4.1 Introduction

EAs are biologically-inspired algorithms for search and optimization that have received much attention as tools applicable to a very wide range of problems. The technique involves artificially 'evolving' a population of solutions to a problem, and this in turn requires the repeated evaluation of a fitness function, which estimates the quality of a candidate solution. In many cases, the fitness function needs to be used thousands or hundreds of thousands of times during the process, and the technique relies on the estimates provided by the fitness function, in order to guide the algorithm efficiently towards good solutions. However, for many real-world problems, several aspects of the fitness function present challenges to the process. To take a pertinent example: if the quality of a candidate solution depends on future events that are difficult to predict, the fitness function will necessarily be noisy and unreliable.

Furthermore, the nature of the fitness may be challenged even for conventional, otherwise accurate fitness functions. This can be the case when the fitness function is available but very computationally expensive to compute, as in, for example, structural design optimisation problems (Ong et al., 2003). The goal in the latter paper, common to much similar work, is to use the real (expensive) fitness function for a limited number of fitness evaluations, in conjunction with an approximate but much faster fitness function, which is used in all other cases. The combination of this model-based fitness, built from a small number of samples of the original function along with the approximate function, is normally called *evolution control*. In other cases, an explicit fitness function may not even exist, as in, for example, art design or music composition. The approach followed here is often to focus on the use of interactive methods, in which one or more humans become part of the process, supplying their opinions, which then take place of the fitness function. (Biles, 1994).

In the concrete context of urban planning, the uncertainties present in objective functions and constraints are significant, and would present a fitness landscape that, intuition suggests, would be extremely difficult for a standard optimization method to navigate successfully. Hence, if the application of EA techniques to solve urban planning problems is to be considered, suitable mechanisms need to be taken on board in order to cope with these difficulties.

However, not all types of uncertainties should be approached in the same way. Jin and Branke (2005) differentiates among four types of uncertainty that can affect the performance of EA techniques. These variants are named: noise, robustness, fitness approximation, and time-varying fitness. We will discuss each of these, but with a special focus on those forms of noise and uncertainty that are most relevant in an Agent-Based System scenario.

Following review of the multiple strategies that have been applied to cope with these types of uncertainty, we will describe and justify the main approach taken in this thesis. As we will see, the latter approach is essentially a carefully constructed approximate fitness function, which is designed to efficiently capture signals of likely future dynamics of the full model simulation. Effectively, our approach builds a statistical model of the agent based system's behaviour, which is able to support rapid approximation of the full fitness function. This approach requires a limited number of prior simulations of the objective function that are averaged and used as an estimate of the real objective value. Then, this procedure allows the application of an evolutionary algorithm to optimise urban growth policies, where the quality of a policy is evaluated within a highly noisy and uncertain environment.

## 4.2 Evolutionary Algorithms under noise and uncertainty

As previously mentioned, EAs can face some issues of applicability when confronted with real-world problems. In the field of optimisation, this strategy, in both single and multiple objective versions, can be characterised as a significant robust method when it has to deal with noisy environments (Büche et al., 2002; Hughes, 2001). Noise means environments where the fitness function is only an approximation of the real fitness function. This advantage is mainly caused by its intrinsic use of a population of solutions to solve the problem under consideration that acts as a filter for noise when the average performance is computed (Arnold and Beyer, 2003).

Sources of noise can be varied. It can be caused by measuring errors related to the representation of sensors and actuators, by the inherent stochasticity of some techniques such as multi-agent simulations, by the propagation of uncertainty in the input data, or by the aggregate behaviour of different factors.

In a noisy environment with a certain degree of randomness, typical of stochastic simulation models, predictability is challenged by the fact that under the same initial conditions and input parameters, results may vary every time they are generated. In Fig. 4.1 this effect is graphically shown. In systems where these variations are inherent and irreducible, data can be represented as a probabilistic distribution.

However, if this randomness can affect the system after the evaluation is performed because the current solution is disturbed, then this specific type of noise is denoted as relating to *Robustness*. Lack of robustness means that the solution may have been evaluated as good, but cannot be used, perhaps due to manufacturing tolerances, which for example directly affects structural design problems (Barthelemy and Haftka, 1993).

In scenarios where noise is present, the selection operator within the EA can

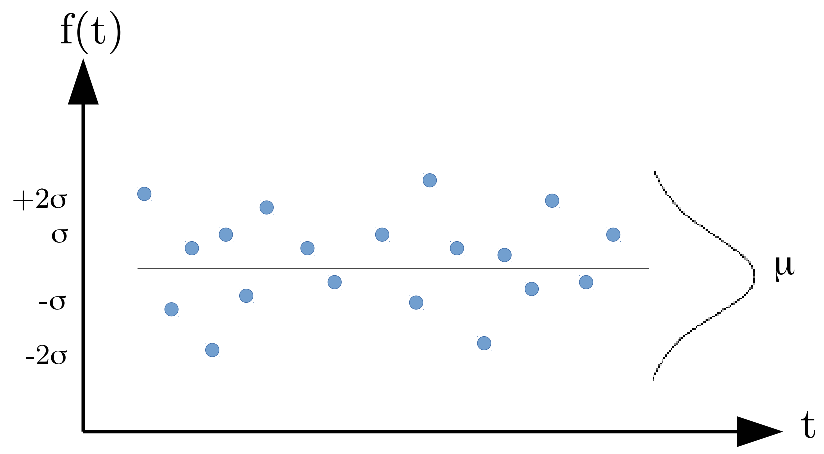



Figure 4.1: Illustration of the variation in fitness values due to noise in a general system. For repeated measurements of the same specific problem, the objective fitness function $f$ changes. In this case, these perturbations are considered to be ruled by a normal distribution.

deliver unstable results and the convergence of the solutions may be adversely affected, propagating inferior solutions (Syberfeldt et al., 2010). In these cases, it is convenient to quantify the probability that the operator generates wrong decisions (Hughes, 2001). This occurs when the fitnesses of the solutions A and B are $f(A) < f(B)$ but their expected distribution values are the contrary $dv(B) < dv(A)$. These distributions can be constructed by performing multiple evaluations for each chromosome, which is very expensive in terms of computational costs. A less demanding approach would be to perform different evaluations in a single random chromosome to estimate the entire distribution, assuming that it can be extended for all the population of solutions. This method will decrease the number of final evaluations, limiting the extra computational resources used.

Another possible risk is the existence of epistemic uncertainty in the system. Galbraith (1973) defines this type of uncertainty in terms of the difference between the amount of information necessary to perform a given task and the amount of information already known. The sources of this type of uncertainty come from scarcity in the amount of experimental data collected, lack of accuracy in the approximations and assumptions selected to simplify the system, significant missing factors not included in the model or even a poor understanding of the processes involved (Oberkampf et al., 2004). To avoid any kind of confusion, from this point irreducible and random uncertainty will be named as noise and epistemic uncertainty will be denoted simply as uncertainty.

In systems characterised by the presence of uncertainty, the definition of the problem under consideration could have a lack of accuracy in objectives values or in the parameters that describe the system. Under these circumstances, a pair of successive evaluations of the same individual solution will retrieve the same objective values and not different ones, as in the previous case. However, these values may not be totally accurate. The complexity in this case arises when two chromosomes are compared. Due to the inaccurate evaluations caused by the uncertainty, solutions can be also misclassified. Generally this uncertainty can be reduced by increasing the knowledge within the system.

Apart from the problem of prioritising the best solutions, the presence of noise or uncertainty in the objectives causes a slower rhythm in the evolution of the population of solutions, since the system may not be able to retain what it learns and fitness does not evolve monotonically, even when elitims is applied (Di Pietro et al., 2004). Hence, taking into account all of these circumstances, for a classical implementation of the evolutionary algorithm, its use and performance have been questioned (Rieser et al., 2011; Wu et al., 2006). From a computational point of view, it is important to mention that epistemic uncertainty is more challenging to cope with than random noise (Oberkampf et al., 2004).

In order to enable the EAs to solve these kind of problems it is necessary to include external tools and mechanisms to support the process. Existing methods that can be applied to evolutionary systems, which work in uncertain and noisy environments include approximation techniques such simplified computational simulations and meta-models (Jin, 2005). However, even if these techniques can aid the EA to be considered suitable for this purpose, an exhaustive search in the literature provided very short number of studies focused on applying EA techniques to Dynamical Planning (DP) or to a SDMP under uncertainty. Instead, this type of problem has been solved traditionally by the family of MDPs. These methods allow the modelling of complex systems based on their preceding state where the objective problem is defined by a predefined set of states and a transition matrix that stores the probabilistic rate of transition among them. The drawback in this case is that such strategies cannot be generally scaled to large problems (Krause et al., 2014).

### 4.2.1 Multi-objective Optimisation

In many practical applications where noise is present, a multi-objective algorithm requires not only to be able to cope with multiple optimisation objectives that can be complex and non-linear, but also with the stochastic noise that is generated as a consequence of uncontrollable variations in the system (Jin and Branke, 2005).

Concretely, in a multi-objective scenario, the system no longer generates two possible outcomes from the comparison between two solutions. Instead there is a triple possible composition that is, $f(A) < f(B)$, $f(A) > f(B)$ and the possibility of non-dominance $f(A) \equiv f(B)$ where the solutions are incomparable, and different decision makers may express a different preference. This extension makes the filtering of noise a harder task. One reason for this increment in complexity is that uncertainty and noise in multi-objective systems changes the nature of the solutions within the Pareto front, which are transformed from points in the search space to hypercubes, see Fig. 4.2.

Noise may alter the dominance relationship between different solutions in such a way that it could be possible that dominated solutions may become non-dominated or vice versa (Tan and Goh, 2008). Consequently, the application of the selection operator may be also misled, eliminating good solutions or reproducing inferior ones. This effect may produce a reduction in the convergence rate and a poor quality set of final solutions (Beyer, 2000; Arnold, 2002; Branke and Schmidt, 2003).

Apart from this aspect, the fitness calculation process may produce outlier solutions whose values are placed at an abnormal distance from the rest of solutions in the search space. In this case, the optimization algorithm might get stuck in one of the solutions which dominates all present solutions (Büche et al., 2001). The appearance of outliers can be caused by insufficient sampling or by the disparity in the distance to the Pareto front among objectives (Babbar et al., 2003).

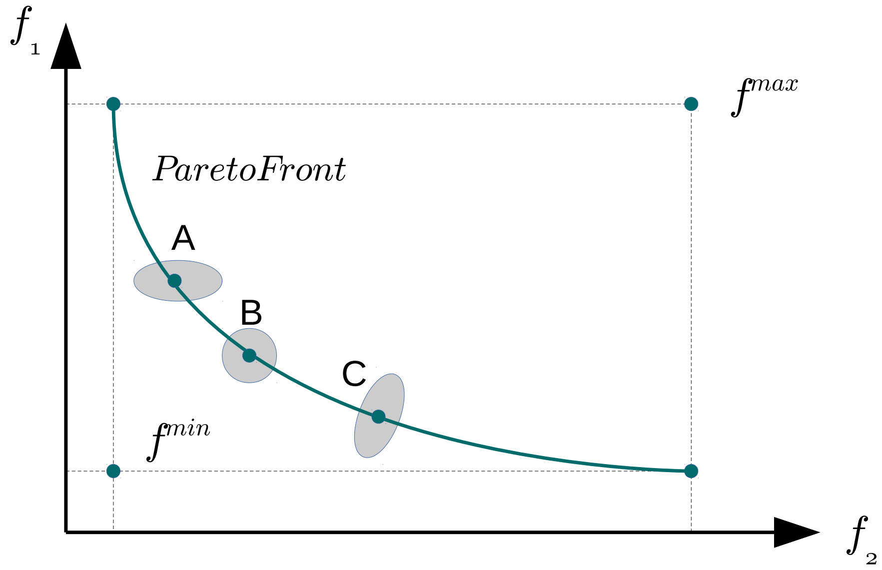



Figure 4.2: Graphical depiction of the difference between the representation of a solution within the Pareto front in scenarios with and without noise. The normal point representation in a standard search space (solutions A, B and C) is transformed in an uncertain environment into a hypercube. This hypercube is represented in this case by grey areas surrounding the point solutions.

Different approaches have been investigated for such multi-objective scenarios. In this regard, a modified Pareto ranking scheme adapted from Goldberg (1989) has been proposed to deal with the presence of noise. There are two major ranking scheme versions that have been studied: one which focuses on probability techniques and another based on clustering methods. The probability-based Pareto ranking schema of Hughes (2001) uses a probabilistic ranking process to take noise into consideration by defining probabilities of dominance between noisy solutions (Teich, 2001). The standard deviation of each evaluation for the entire population of solutions can be used to correct the noise. In this technique, the probabilistic rank of an individual is calculated by the sum of the probabilities of those solutions that this chromosome dominates. Finally, in the clustering variant (Babbar et al., 2003), the Pareto front is formed by the best found solutions plus solutions that belong to their neighbourhood. The neighbourhood calculation takes into account a user-defined restriction factor and the standard deviation for each objective.

Additionally Büche et al. (2001) proposed a modification of the $(\mu, \kappa, \lambda)$ algorithm (Bäck et al., 1991) to minimise the effect of noise and outliers. The Domination

Dependent Lifetime (DDL) method assigns a maximal lifetime $\kappa$ to each individual based on the number of solutions it dominates, in such a way that the lifetime value $k$ will be shorter if the number of chromosomes in the population is large. This feature contrasts with the effect of elitism which may preserve solutions for an infinite amount of time by limiting the impact of inferior new individual solutions. However, to prevent the elimination of good solutions, the approach is complemented with a mechanism that allows the re-evaluation at the end of the lifetime of the expired solutions. If these solutions are good enough they will be added again to the pool of solutions with new objective values resulted from a new re-evaluation. This procedure, however, will only replace previous good solutions with other noisy samples.

## 4.3 Fitness Approximation

In an uncertain and noisy context the fitness function, which is evaluated by means of statistical, conceptual or physical simulations, is normally the most computationally intensive element of the given application (Nicklow et al., 2009). This high computational requirement has led to the development of approximative alternatives to alleviate the corresponding cost. Surrogate models, also known as metamodels, are used to replace costly simulation models (Barton, 1994). These models lead to a more efficient exploration and exploitation of the search space.

There are different types of techniques that can be used as a modelling tool for function approximation, aimed at replacing computationally intensive models. This is the case of the use of ANNs (Funahashi, 1989; Hornik et al., 1989; Chambers and Mount-Campbell, 2002), Least Squares Support Vector Machine (LSSVM) (Van Gestel et al., 2001; Wan et al., 2005) or Kriging interpolation (Ratle, 1999) or response surface methods which uses least-squares regression based on low-order polynomials. See (Giunta and Watson, 1998; Simpson, 2000) for an extensive comparison between these techniques. The critical issue in this approach is to find a good quality approximation strategy in such a way that the behaviour of this approximation is similar enough to the original model. Otherwise, the final system could experience a severe negative impact when errors are evaluated (Jin et al., 2002). Each method should deal with the avoidance of over- and under-fitting, to capture enough complex casual relationships in highly dimensional problems.

Formally, if the real representation of the problem under consideration is the following:

$$y = f(x) \tag{4.1}$$

the corresponding surrogate model is an approximation of the form:

$$\hat{y} = \hat{f}(x) \tag{4.2}$$

in such a way that $y = \hat{y} + \mathcal{E}$, where $\mathcal{E}$ is an additive error term.

Surrogate models can be built by an iterative online process, using only the data points that belong to the local neighbourhood of the point of interest (Ong et al., 2003) or taking information acquired in the past. The latter is called Memory Based Fitness Estimation (MBFE) (Sano and Kita, 2002). In this strategy, a stochastic model that is capable of gathering data about the uncertainty of the fitness function is used. Sampled values of the fitness function are acquired and stored in a memory-based system forming the search history of the problem. Finally, when the fitness value of an individual needs to be estimated, statistical techniques over this repository are used of information to retrieve the corresponding value. This approach was selected for this thesis as a method to deal with uncertainty.

## 4.4 Sampling Fitness Function

The manner in which noise influences the fitness value can vary. In additive noise, additional values are randomly added to or subtracted from the real fitness value. Formally, this type of additive fitness function can be defined as such: if $\rho_i$ is the fitness function that is defined based on a determined configuration of the problem for a determined chromosome $i$, then the noisy fitness function $\rho_i'$ can be described as:

$$\rho_i' = \rho_i + rnd[N(\mu, \sigma^2)] \tag{4.3}$$

where $\rho$ is the noise–free fitness function and $[N(\mu, \sigma^2)]$ denotes the assumption that the noise can be approximated by a normal distributed noise component added in each evaluation. As a facilitator assumption which is reasonable for many domains, it can be also assumed that noise is unbiased $[N(0, \sigma_N^2)]$. Additionally, the uncertainty set $U$ of $\rho_i$ can be defined as:

$$U(\rho_i) = \{\xi \in \mathbb{R}^n : \rho_i - \triangle \leq \xi \leq \rho_i + \triangle\} \tag{4.4}$$

where $\triangle = (\triangle_1, \triangle_2 \cdot \triangle_n)^T \in \mathbb{R}^n$ is the aggregate uncertainty, $\xi$ is the neighbourhood created by the uncertainty and $n$ is the dimension of the decision space. The size of $\xi$ is defined by the internal boundaries $[\rho_i - \triangle, \rho_i + \triangle]$.

The use of a normal distribution is very common (Büche et al., 2002), however the nature of the source of noise can be characterised by other types of distribution. Based on the central limit theorem, the sampling mechanism reduces the amount of noise by calculating the mean of multiple function evaluations.

$$\rho_{i,n}^* = \frac{1}{n} \sum_{j=1}^{n} \rho_{i,j}' \tag{4.5}$$

where $\rho_{i,j}'$ is the sampling realisation number $j$ of the individual $i$ and $\rho_{i,n}^*$ is the distribution of $\rho$ resulting from the mean of $n$ samples of $\rho'$. As the sample size $n$ is increased, the standard deviation is reduced.

In order to generate the analytical values required to define the fitness by approximation, four basic strategies have been introduced, namely explicit averaging, implicit averaging, fitness inheritance and selection modification (Jin and Branke, 2005). All methods assume that the search space is characterised by a known and homogeneous noise distribution, most commonly a uniform or a normal distribution type. It is also considered that an estimation of its magnitude is possible to calculate (Bui et al., 2005). However, these assumptions limit the effectiveness of the selected approach due to, in general, the effect of noise is not spread homogeneously over the search space and the absence of knowledge regarding the level of noise are the most common characteristics of real-world problems.

Explicit averaging, also called static resampling, was introduced by Miller (1997) and it is the most commonly used method for coping with noise. The strategy consists of generating a determined number of times the sampling of the objectives, followed by the averaging of the generated values (Fitzpatrick and Grefenstette, 1988). In a sample size of $n$, this operation allows a proportional reduction of the variance by a factor of $\sqrt{n}$. Additionally it also means scaling up the required computational effort used by a factor of $n$ (Jin and Branke, 2005). To avoid extra evaluations, the fitness from the neighbourhood can be used (Branke, 1998; Sano and Kita, 2002).

Another possible approach is to apply a statistical model constructed beforehand with historical data to model the fitness using techniques such as local regression and adaptation (Branke et al., 2001). Sampling is a popular method to reduce noise and estimate unknown information. If the approximate model is generated by an offline training process before the optimisation is run, it is common to use Monte Carlo techniques to generate these samples. However, if evaluating the fitness function is significantly time-consuming, this strategy may not be viable.

In the implicit averaging method, on the other hand, sample size is defined as an inverse function of the population size (Fitzpatrick and Grefenstette, 1988). The idea behind this interpretation consists of that in systems defined with a large population of solutions, it is very common for there to be numerous chromosomes that are very similar to each other. The frequent evaluation of these related areas of the search space reduces the noise.

Bui et al. (2005) introduced a technique to solve this problem based on the idea of fitness inheritance. They proposed that the offspring created in each generation additionally inherits two variables from its parents: $\mu$ that represents the mean of the objective value and $\sigma$ that corresponds to the standard deviation. These variables will control whether a new resampling is required or not. The resampling operation consists of calculating the new fitness by performing a predefined number of evaluations where the final values $\mu$ and $\sigma$ correspond to the mean and the standard deviation of these evaluations. When a new child is evaluated, a resampling is only required if its objective values fall outside the confidence interval. Otherwise, the inherited fitness is assigned. Consequently the evaluation of solutions characterised with higher noise will result in larger standard deviation values which facilitate the fitness acceptance in its children.

Finally, the modification of the selection operator is another method investigated to cope with the noise when the fitness reevaluation is too costly. Teich (2001) defined a selection and ranking procedure that takes into account some conditions like the probability of dominance to compensate the noise. Another similar strategy uses a threshold value when fitnesses are evaluated to overtake the effect of the hypercubes in a multi-objective scenario (Markon et al., 2001). A useful survey of such mechanisms appears in Jin and Branke (2005), while a brief update of the state of the art was shown in (Qian et al., 2013).

### 4.4.1 Monte Carlo simulation

Monte Carlo simulation is a technique that was developed in the 40s by Metropolis and Ulam (1949). Since then it became a widely used and effective tool for those problems whose analytical solutions do not exist or have a high level of complexity to be easily obtained. By means of random sampling, the strategy allows the study of the properties of random-nature systems when analytical solutions are not easily available. To recreate properly the desired dynamics and patterns of the studied system into the model, it is common to use real information gathered from this objective system. However, in some cases the information collected in this way has not enough quality or cannot be easily measured and structured as a probabilistic distribution. It can be also possible that even if this information exists, its application in a large stochastic model could be a very challenging task (Huang et al., 1992; Li and Huang, 2009; Lv et al., 2010).

The number of draws used within the sampling should be defined according to the level of noise and uncertainty which characterises the search space of the problem.

In general, very noisy scenarios will require extra samples to come up with the same level of robustness than in more deterministic search spaces (Syberfeldt et al., 2010). However, each new sample will increase the computational effort required for generating a single evaluation. If Monte Carlo techniques are used along with an EA approach, an alternative option to manage the noise is to increase the number of individuals that forms the population of solutions (Harik et al., 1999; Miller, 1997). However, it could be hard to know a priori, the most efficient size for the population, because this aspect depends on several factors including the level of noise, formulation and problem-specific parameters (Harik et al., 1999; Miller, 1997). At this point, there is controversy surrounding the trade-off created between the role that these two factors plays in decreasing the level of noise (Cantú-Paz, 2004). Fitzpatrick and Grefenstette (1988) and Arnold and Beyer (2001) highlighted the size of the population to increase the robustness against uncertainty over the sample size. Meanwhile, Beyer (1995) and Hammel and Bäck (1994) favoured the sample size instead. However, these conclusions strongly depend on the definition of the problem. In this regard, these authors state that for the $(\mu/\rho, \lambda)$-ES (Beyer et al., 2002) an increment in the population of solutions is preferable when the parameter called truncate ratio $\mu/\lambda$ is calibrated appropriately. However for $(1, \lambda)$-ES (Oyman et al., 2000), averaging over multiple samples is the best option. In this type of problem, $\mu$ refers to the size of the parent population, $\rho$ is the number of parents involved in the mating of a single offspring, and $\alpha$ denotes the number of offspring per generation.

Under these circumstances, it is challenging to exactly know before the algorithm is empirically tested, the reliability or level of robustness of a determined formulation. For single objective problems, Miller and Goldberg (1996) inferred a lower bound of the optimal sample size and suggested that, in a system with uncertain parameters, the EA solutions only require the generation of a small number of samples. They stated that a limited number of Monte Carlo draws, that can range from 5 to 20 per population member, should be enough to compute their average fitness. This assumption is based on the idea that in EAs new samples are included in the population in each generation by the application of elitist operators that highlight good solutions. As a direct consequence of this mechanism, this process will have the side effect of implicitly increasing the number of Monte Carlo realizations.

A proposed extensions of this approach is the introduction of an operator that limits the age of the solution, which control the survival of fit members. By the use of this element it is possible to further reduce the number of Monte Carlo draws in the fitness evaluation (Hilton and Culver, 2005; Kapelan et al., 2005; Wu et al., 2006).

## 4.5 Description of the problem

In a dynamical location-allocation problem where a set of urban green areas have to be allocated during a determined period of time subject to some constraints, the major objective to achieve can be defined as satisfying the needs of the population. This satisfaction can be measured in terms of the distance from households to these green areas, since access and frequency of use are mainly determined by the distance between the dwelling to the park (Giles-Corti et al., 2005). The availability of parks at a close distance provides varied types of beneficial services and amenities that these green facilities offer to the population from different perspectives such as aesthetic, physical, social and environmental (Chiesura, 2004; Bowler et al., 2010b,a). The model does not include the option of using any means of transport to arrive at the green area in the case it is far from their residence. The search can be extended to cover other objectives like environmental protectionism, level of connection between areas and profitability among others. This conflicting set of goals, more typical of real-world problems, arguably needs multi-objective techniques to be appropriately solved.

A 'policy', in this context, amounts to the city authorities' planned schedule for protection of a specific set of green spaces maximising the objectives selected in both, short and long-term. In these scenarios, financial resources are normally accessible periodically. However, because of governmental purchase decisions are subject to the availability of parcels of land and this factor changes over the time, the capacity to provide green spaces during the construction of new urban developments in cities is normally limited. Hence, careful planning studies should be carried out in advance. Computational optimisation techniques can be applied to the search for an ideal policy in the face of budget constraints. It is worth mentioning that the budget may normally be quantitatively much lower than current prices of the patches of land that are significant for the new areas under construction. Additionally, since land prices generally increase with time because of multiple factors, including the rise in the demand of these spaces, scarcity of available land and other related economical factors, current acquisition policies should take into consideration not only the present status of the system but also a reliable projection of future necessities. However, dealing with future conditions implies that we need to cope with epistemic uncertainty due to a lack of knowledge about the future.

In this regard, there is much active research in designing long-term feasible public open space plans, whereby researchers interested in urban planning and sustainability have investigated a range of agent-based systems and similar mechanisms to explore the consequences of different green-space allocation strategies (Parker et al., 2003; Sasaki and Box, 2003; Sanders et al., 1997).

In general, the application of modelling techniques is another element that aggregates epistemic uncertainty, mostly inherited from the selection of the model and the subsequent structural changes required to adequate the system to the considered problem. Apart from that, due to the use of a CA-ABM framework that, this work applies as a modelling technique, it should be noted that these concrete technologies implicitly add noise to the system into consideration. Consequently, the applied EA algorithm should be robust enough to be able to cope with both factors at the same time, in order to provide valid and usable results.

In such a context, a method that effectively obtains a model-driven approximation of the simulation to lead the evolutionary algorithm towards policies that yield much improved satisfaction levels than unoptimised policies is investigated. By collecting the knowledge needed for optimisation by the simulation process, we can consider using a rapidly accelerated model of the agent-based simulation in place of the real knowledge. This requires a limited number of prior simulations of the agent-based urban growth system, and then allows the use of an evolutionary algorithm to optimise urban growth policies. This strategy is based on the fact that, even if approximate models do not have the capability of creating new information, they can gather useful information from the history of the optimisation and prevent its loss (Ratle, 1999).

### 4.5.1 Statistical Data Generation for Sampling

Specifically, the EA algorithm requires to receive, as input parameters, concrete information about the scenario in which the algorithm is operating. However, if some of these elements that characterise such a system are totally or partially unknown, then an external mechanism should generate the information about these uncertainties. The way in which this is performed will have a significant impact on the feasibility of the proposed solution.

In this concrete problem, this lack of information occurs due to the complexity resulting from the multiple interactions between the different processes involved in the stochastic growth of the urban development model, that cannot be foreseen beforehand. The relationships among these factors can lead to the development of a varied range of future scenarios.

Data sampling techniques used to create the approximation model could cover both online and offline learning. The online approach collects and updates the system during the optimisation, meanwhile the offline version performs a training process beforehand, incorporating prior knowledge. One of the advantages of using non-adaptive techniques is that, once it is demonstrated that the sample size describes a function that is Pareto-optimal in relation with the speed and the accuracy of the algorithm (Srinivas and Deb, 1994), the offline procedure permits the system to focus mainly on the accuracy factor. This is based on the idea that the offline process can invest all the time it is required to generate the set of samples to achieve the best accuracy since the speed is a non-significant factor.

If the sample set will be generated by an offline process, the number of samples collected will be decided beforehand and the statistical model will remain constant for the entire optimisation process. Other approaches define a variable number of samples for different individuals or for different phases of the optimisation process. Aizawa and Wah (1994) focus on minimising the expected estimation error, sampling from the best individuals in the population and Branke and Schmidt (2003) take further samples from individuals included in the mating pool which where selected using a Tournament Selection technique.

After that, the required information that the EA will use is sampled 20 times each for every factor analysed, to form an initial estimate of the amount of noise. This number was selected because it was empirically calculated that further samples do not add new significant information to the set. The factors analysed are the amount and distribution of the population, prices of rural land and urbanised areas. Based on the collected samples, the mean and sample standard deviation are calculated. This process is equal for single and multiple-objective scenarios. The difference when passing from a single to a multi-objective paradigm is that the amount of data gathered can be different if there is enough uncertainty involved in any of the new objectives considered.

Miller and Goldberg (1996) stated that this sampling technique makes the EA algorithm highly reliable, performing even better without extensive sampling. The reduced number of samples has a computational advantage compared to other approaches that require the generation of a much larger number of samples to achieve good results (Murray and Church, 1995).

### 4.5.2 Sources of Uncertainty

In this urban scenario, the uncertainties and variabilities arise mainly from the following sources: (1) multiple choices for transforming rural areas into green parcels; (2) uncertainty about urban growth evolution; (3) lack of knowledge of the future total population and its distribution; (4) uncertainty regarding urban and rural land prices dynamics; (5) land resource availability.

Other works investigate the influence of a non-deterministic budget in the planning policy, where the availability of future resources is unknown in advance (Golovin et al., 2011).

Some factors which actively contribute to the increment of the level of uncertainty are further described below:

#### 4.5.2.1 Urban Property Prices & Green Areas

By using hedonic prices and contingent analysis techniques, it has been widely demonstrated and analysed in the literature that green spaces exert a direct positive influence on the prices of the surrounding residences (Tyrväinen and Miettinen, 2000; Thorsnes, 2002). This aspect is included in the model as the general desire to live close to one of these areas and it is represented by the agent's acceptance to pay more for any of these specific locations. The inclusion of this personal inclination in the utility function used by the agents causes a significant growth in the demand of these areas and subsequently in their price. This non-homogeneous population concentration within the city affects the normal rhythm of urban spatial spread which is sped up in these areas.

#### 4.5.2.2 Ecological Degradation & Non-Urban Prices

From the perspective of the set of non-urban cells, one of the two main parameters which involves a high level of uncertainty is the relationship created between the non-urban price dynamics and the cells' ecological value. Recalling two stochastic characteristics of the model, firstly that ecological changes in a specific area of the lattice influences its neighbourhood, spreading in all directions, (see the behaviour of the *bioValue* Algorithm 3.3 on page 111) and secondly that the urban growth process depicted by the expansion of its boundaries to new peri-urban areas provokes an ecological degradation of the entire area (see Formula 3.18). As a consequence of both factors, the ecological value of the open areas declines irrevocably as the boundaries of the city expand.

Furthermore, this non-deterministic ecological degradation process affects in turn the prices of non-urban areas. In formula 3.17, that controls the price of these underdeveloped areas, one of its terms, called $P_{base}$, is a constant factor whose value varies if the cell is classified as agricultural or forest. In the model, this classification is not static, instead it depends on the ecological value of the cell under consideration that varies through the time.

#### 4.5.2.3 Urban Growth

The underlying process of urbanisation is in nature partially random and mainly determined by three factors:

- The transition rules of the cells defined within the CA are based mostly on a set of preselected probabilities. See algorithm 3.2.
- The amount of the new population that wants to settle down in the city which mainly controls the transformation speed of peri-urban areas into new urban cells.
- The amount of green areas located within the limits of the city that causes an attraction effect over the demand of the surrounding areas from new potential inhabitants in comparison with other possible candidate areas.

However, the relationship created between urban and green areas is bidirectional. Knowledge of the urbanisation process is crucial for the selection of green areas because the set of candidate cells to be protected is restricted to the cells that are underdeveloped and, hence, a cell that is already urbanised cannot be transformed into a park. The system, subject to this restriction, needs to know the complete state of the cells in each time step in order to properly select and protect non-urban cells in advanced.

#### 4.5.2.4 Flows of Population

Another significant characteristic of the model is that the city is a non-closed-system. This means that there is an external flow of new population coming from migration as well as new offspring resulting from the current settled population. The dynamics of these flows and the density of each future neighbourhood are not fixed and easily predictable even if there exists a general preference to live close to the city centre in line with the Alonso (1964)'s model.

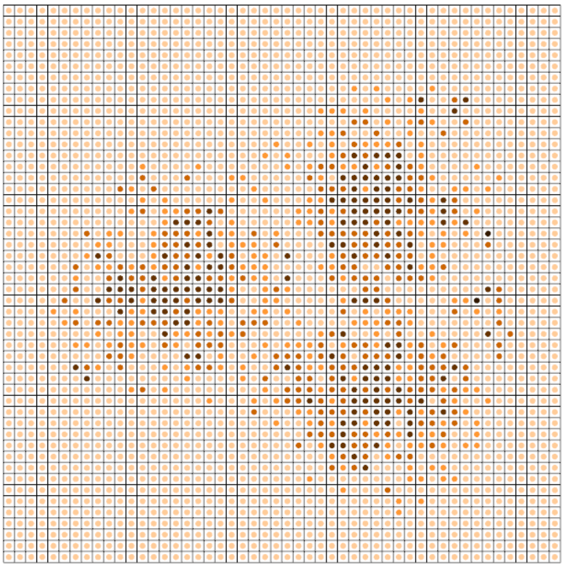
Consequently we are not able to know in advanced the percentage of population directly affected by a determined location of a new green area and hence, the final satisfaction achieved by a determined configuration of green spaces.

### 4.5.3 Model Application

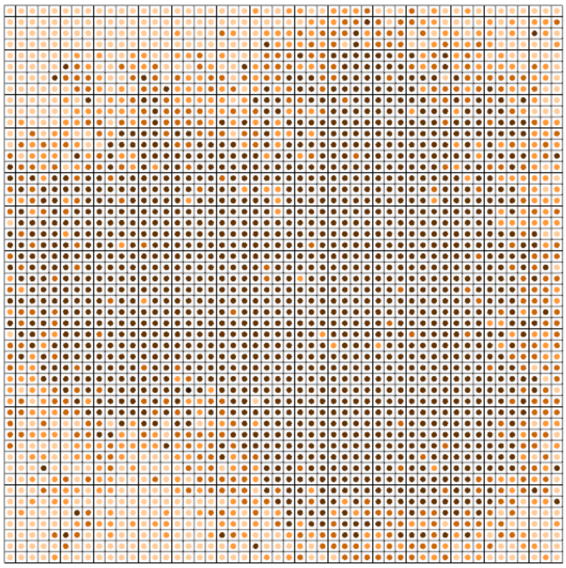
As was seen in Eq. (5.12), page 168, independently of the nature of the planning (online or offline) or the number of objectives considered, if the satisfaction of the population is included in the calculation of the fitness, the expected population distribution needs to be inferred. Considering that it is impossible to know in advance the future spatial configuration of the population within a city and that such a city is a system with complex spatial and temporal dynamics (Jacobs, 1961) then, in order to be able to measure the distance from each agent to the closest green area, external tools are necessary to incorporate into the EA. These tools will allow to assess the appropriateness of the defined set of candidate solutions.

To assign values to this uncertain parameter, a Monte Carlo sampling strategy is adopted to recreate a plausible population distribution that is used to compute the fitness. This method can be seen as an offline sampling fitness function (Smalley et al., 2000).

In the model, the fitness sampling is implemented using the generation of a sample set of equally likely realisations that captures the spatial population dynamics through time (Rada-Vilela et al., 2014). The source of these realisations is an equivalent version of the same model that is statistically similar to the actual site without the inclusion of green externalities. Excluding this type of externalities implies to remove possible non-linearities resulted from the relationships created between urban prices and green areas (Wu and Plantinga, 2003), residential preferences and parks (del Saz Salazar and García, 2007) and the ecological degradation and the dynamics of the non-urban prices, but keeping the general evolution of the urban growth system. These perturbations are particular to each individual realisation. In this thesis, this



(a) Lattice with the information used to calculate the fitness approximation for the time step 300.



(b) Lattice with the information used to calculate the fitness approximation for the time step 600.

Figure 4.3: Representation of two different time-steps of the simulation, where the amount and distribution of the agents used to construct the approximative fitness function is visually depicted. The population distribution is represented in a range of red colours, where darker tones represent, in relative terms, the most crowded areas of the system and lighter colours unpopulated zones.

simpler version of the urban model will be named a *surrogate model* and it will only be used before the optimisation process starts.

Matrices are formed by collecting samples of the values of the probable quantification of the analysed dynamic, in each of the cells of the grid, for every time step of the simulation. This dynamic will correspond, in the case of the fitness function, to the position of each agent. Matrices will store the accumulative value of running the surrogate model $n$ times, along with the number of runs. A single division will be done when each value is finally used, in order to avoid any loss of precision in the final value. The value of $n$ will not be less than 20 in order to properly mimic the richness of the dynamic. A set of 600 matrices, one per time step are created for each data type required in the problem into consideration as it is shown in Fig 4.4. Once this information is available, they will be used as inputs for the different optimisation algorithms. A visual representation of these matrices is depicted in Fig. 4.3a and 4.3b, where the range of values are normalised and assigned to a range of colors with similar tonality.

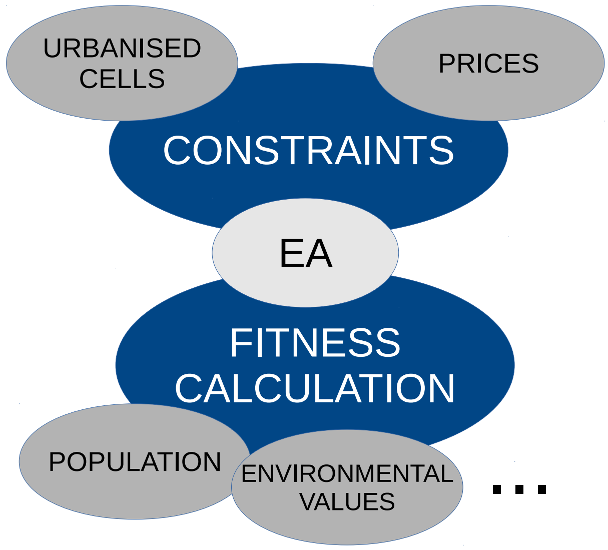



Figure 4.4: Sources of data collected to support the EA optimisation process in the urban environment.

As such, the fitness function is estimated for each considered time step by using the information gathered by Monte Carlo sampling. The noise of each chromosome $X^*$ is reduced by calculating the fitness function of individuals which belong to a similar search space which was previously evaluated in an offline process. This approach increases the estimation accuracy without the necessity of performing extra online evaluations.

The same gathering method has been successfully applied in other fields including dialogue systems (Rieser and Lemon, 2011), in environmental studies (Kennedy et al., 2006) or emulators for managing uncertainty in complex models, such as the Multilayer Urban Canocopy Model (MUCM) that simulates the features of the urban climate (Kondo and Liu, 1998; Crucifix and Rougier, 2009). By means of such method, these systems are capable of gathering the required data by an offline sampling mechanism.

Different sources of data are collected to support the EA optimisation process, see Fig. 4.4. These data can be divided into two groups according to the role they play within the urban model. They can serve as a constraint when the rural area is selected or be involved in the calculation of the fitness. In the single objective version of the model, two sources of data, the prices of the non-urban areas and how the urbanisation spreads over the grid, are constraints and the population density and its distribution are used in the calculation of the fitness. In a multi-objective version of the framework, additional factors could be included in the calculation of the fitness to cover other aspects of the optimisation process such as environmental values that assess the ecological richness of a given patch of land.

The design an nature of the optimisation algorithm determine the amount of data needed. As already mentioned, in the discrete model studied in this work the selection of green locations is performed sequentially in each time-step which is economically limited by the current *budget* of aggregate nature, and the configuration of the lattice in the precise moment the decision is made. These restrictive factors do not include information regarding the satisfaction achieved by the population. If only the current needs of the population should be taken into account, the information related to satisfaction can be collected in the same way as the constraints. However, if future long-term conditions are taken into account to develop present policies, a different approach should be followed.

### 4.5.4 Significance of Results

To test the appropriateness of the method, the collected population distribution was compared with the data gathered from an instantiation of the model used for testing. The collection of the test data is performed at the end of the simulation, once the optimisation terminated, using the same module responsible for testing the offline policy and checking the feasibility of the policy. The concept behind this is to use this instantiation of the testing model as a 'real' data, that is the real outcome that happens when the green policy is in use. The final results generated in this 'real' execution depend on both, the intrinsic effect of random noise (introduced in 4.2) and the values of the parameters involved in the study. The significance of the results will be calculated using the distribution gathered from the Monte Carlo sampling and the same distribution generated by the 'real' outcome.

In order to calculate the degree of similarity between these two distributions, correlation analysis was used. By means of canonical correlation techniques, it is possible to estimate a symmetric measurement of the congruence of two matrices (Ramsay et al., 1984). A canonical study implies the analysis of variables that are not directly observed but that represent the ones that are directly observed. In this concrete case, Pearson's linear correlation matrix is selected to measure the strength of the association between the two matrices: the Monte Carlo pregenerated matrix $M'$ with the matrix composed by the 'real' population distribution $M$.

Since the goal of the problem is to study the system during a period of time, instead of having two matrices corresponding to the values of these two distributions, the problem is represented by a series of matrices, one per time step. Then, the first step in the calculation will transform this set of matrices into a single matrix, combining all of them within a unique data structure. The result is two matrices, one with the sampling data an another with the 'real' data, both with dimensions $600 \times 2500$, resulting from vectorising the grid of $(50 \times 50)$ values that represents the population within the city in each time step. $T = 600$ is the time horizon considered and each column will store the value of a cell of the lattice in the time step represented by its row. Every value of $M$ and $M'$ is based on averages of 20 different observations (simulation runs) for each time step considered. The resulting correlation matrix calculated from the real population distribution $(M)$ and the simulated sampling distribution $(M')$ is 0.7634. Fig. 4.5 shows a visual representation of the data, where the values of the two distributions are plotted along two axes.

This value shows a strong correlation between both sources of data. To validate this conclusion, the matrix of p-values for testing the hypothesis of no correlation between both matrices was calculated. P-values is a statistical method of testing hypothesis, measuring the strength of the association between the two matrices.

In this case, the p-value is equal to 0.0000, which means that it can be concluded that the assumption that the correlation is due to chance can be rejected. Consequently, the use of the approximate fitness function within the EA algorithm makes the final results reasonably reliable in comparison with real fitness evaluations, if the system is supported by a robust and consistent urban model. It is also concluded in other comparable studies that an evolutionary method performs even better without extensive sampling (Miller and Goldberg, 1996; Vallejo et al., 2013). The reduced number of samples is a computational advantage compared to other approaches that require the generation of a much larger number of samples to achieve good results (Murray and Church, 1995).

Since our noisy fitness function uses sources of information generated by Monte Carlo sampling, the fitness function can be seen as a type of sampling fitness function (Smalley et al., 2000) even if this sampling procedure is done offline.

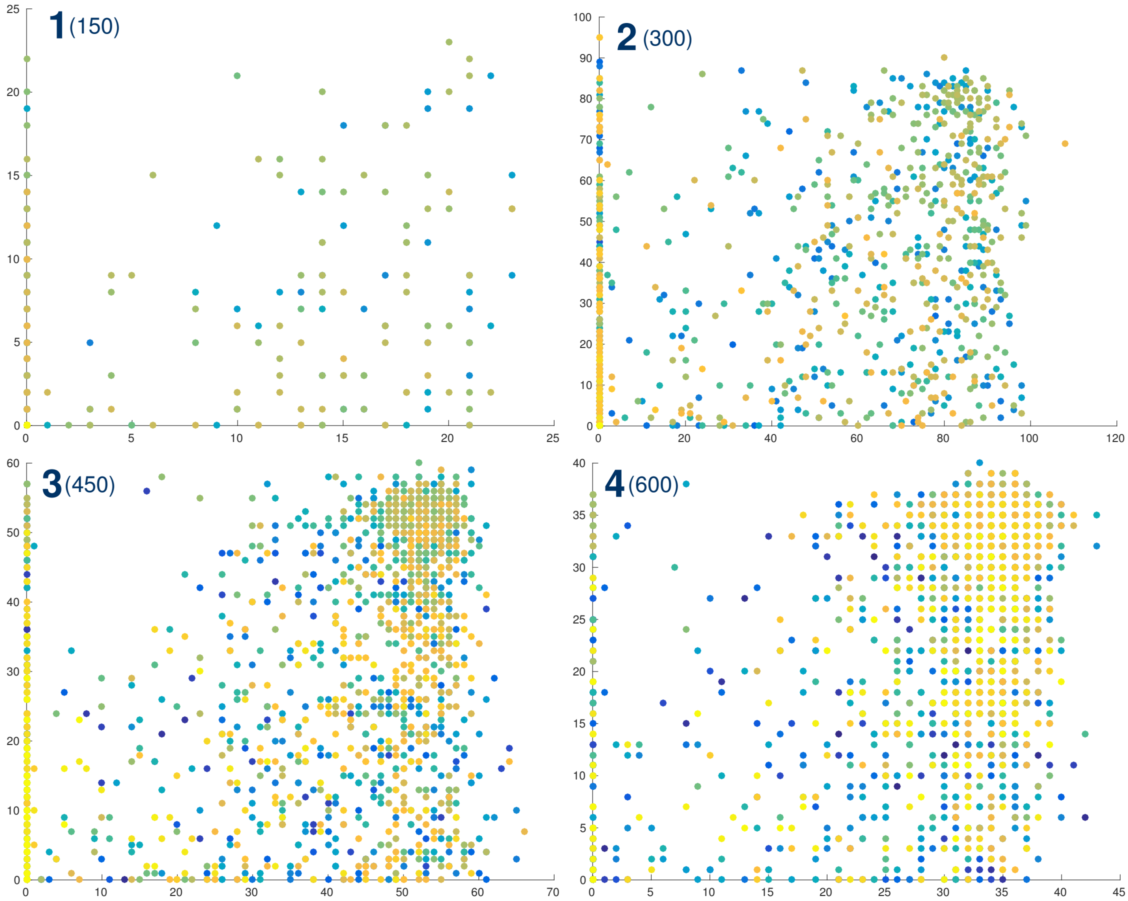



Figure 4.5: The four scatterplot figures shown here is a bi-variable representation of the data gathered in the model. The sampling and real distribution are plotted one against the other for four time steps of the system (150, 300, 450, 600). X-axis represents the 'real' data and Y-axis the one gathered by Monte Carlo sampling. A range of colour gradients represents the position of the point in the grid. The darkest blue corresponds to the cell(1,1) and lighter yellow the cell(600,600).

## 4.6 Conclusions

In this chapter, we have looked at a subset of difficulties that EA techniques may face when they are used in real-world problems with noise and epistemic uncertainty.

Different kinds of uncertainty were described, along with approaches that have been used to handle them in the optimization literature, considering both single and multiple objective optimization. Afterwards, we returned to the central optimization problems addressed in this thesis, and discussed them with a focus on the sources and types of uncertainty inherent in them. We then introduced and discussed the method designed in this thesis to enable the optimization methods to cope effectively despite these uncertainties. In short, this is done by means of Monte Carlo sampling used to inform a model of the agent-based system, which is able to deliver efficient information about the future consequences of planning decisions.

In the next two chapters, the proposed optimisation method is applied, using the techniques discussed here, to a typical urban growth simulation, in which the overall goal is to find policies that maximise the 'satisfaction' of the residents. The strategy is applied to the problem with different configurations of the EA. In its single-objective version, an offline EA methodology is applied within a set of different scenarios where multiple levels of complexity are considered. This is followed by the application of the same techniques cast into a more efficient 'online' approach, where a series of optimizations each makes a single planning decision.

# Chapter 5

# Optimisation of green spaces allocation

## 5.1 Introduction

One of the most common themes in the field of urban planning is to study the dynamics involved in urban growth, which is linked with the relative distribution of urbanised areas, industrial facilities and green spaces. The final configuration of an urban environment, and how that depends on the broad strategies put in place for managing the evolution of different types of land-use, have a significant impact on quality-of-life issues (Robinson et al., 2012).

One of these key land-use types is urban green spaces. Green spaces play a crucial role in the provision of healthy environments in densely populated areas (Groenewegen et al., 2006). In this regard, one of the most urgent research issues within the field of urban sustainability is the study of mechanisms that can mitigate the ecological degradation that is linked with modern urban expansion, a phenomenon called urban sprawl (Mills, 1981), while at the same time ensuring a proper level of provision for the city's population. To meet the population's present and future needs, public direct land acquisition is one possible tool that can be used (Acharya, 1987; Heimbürger, 1976).

However, this land acquisition process is far more complex than, for example, arbitrarily choosing a random number of patches of land to be transformed into green areas. Planners must consider a variety of things that tend to significantly restrict the space of possible acquisitions. Among these are: influential stakeholders who may have conflicting goals, limited budgets and other external factors that may narrow the possible set of options. In some cases, due to the intrinsic level of complexity of these decisions, experts may delegate their spatial assignment choices to computational models, which can be tasked with applying exact optimisation techniques in order to find the optimal distribution of facilities. However, in real-world situations, the complexity and the high computational time that such an exhaustive search could entail, will typically mean that these methods cannot be used at all. Hence, normally, other types of optimisation strategies such as metaheuristics are used instead.

Another factor that complicates the allocation process is the uncertainty linked with long-term plans. In this regard, when a sequential set of decisions has to be taken in advance with no information about the state of the environment at these times, restrictive conditions and unexpected dynamics can easily arise. Comprehensive long-term plans need to be designed under a handleable range of unpredictable future scenarios (Rounsevell and Metzger, 2010). The crucial issue here is how to figure out which are the most likely scenarios.

In this chapter we address the problem of solving dynamically a sequential decision making process in which new financial resources become available over time and where decisions depend on choices taken previously and on most probable future conditions. Concretely, we apply an evolutionary algorithm to solve this problem, analysing how the level of complexity of the selected configuration of the urban area can affect the performance of the final solutions, and how, under certain circumstances, a simple greedy heuristic can find an optimal solution more easily than the proposed more elaborate method.

## 5.2 Problem Definition

The general objective of the problem consists of designing an offline planning process which leads us to find the optimal subset of green spaces out of a set $V = \{1...n\}$

of locations along with the corresponding time schedule of each of the purchase decisions. The offline nature of the selected planning approach means that the policy construction step will be done entirely before the plan is executed.

Each element within the set of different purchasing plans $P$ includes a set of parcels of land $T \subset V$ located within a given geographic area close to or within a city, that are intended to be acquired for conservation and/or social purposes. For the sake of simplicity each patch of land has a homogeneous size and shape. Each of them is considered an independent unit and no clustering techniques to group them are implemented in the model. We also assume that once an open area is selected and transformed into a park, it remains protected from urbanisation until the end of the planning horizon.

Formally a certain purchasing plan $p \in P$ is depicted by a set of parcels of land $\tau \in T$ and a purchase schedule $\psi \in \Psi$ which can be defined as a mapping from the parcels contained in $T$ to a series of purchase times in $\psi = \{0, 1, \cdots, H\}$, where $H$ is the maximum time horizon considered in the plan.

Commonly each candidate patch of land $\tau_i$ has associated with it its own cost $c_t(\tau_i)$ which covers the acquisition and the restoring/transforming process from a rural patch of land into a green space. This cost is calculated at time $t$, which is the moment when the area is purchased, and it is defined for this specific problem as a monetary cost. The value of this cost, which is always strictly positive, is not static and could vary over time. After that step, no further maintenance costs over the area are considered. Under these circumstances, every purchase schedule included in $\Psi$ is linked with a corresponding non-decreasing cost function $C_\Psi$ for the entire conservation plan. This 2-tuple $(\Psi, C_\Psi)$ describes the purchase history in relation to the accumulated cost of the land in such a way that:

$$C_\Psi(H) = \sum_{t=0}^{H} c_t(\tau) \tag{5.1}$$

where $t_0$ is the starting point in time in which acquisitions can be done and $H$ is the maximum time covered in the planning process. Since budgets that can cope with single purchasing transactions involving large extensions of land are normally rather unlikely in real-world scenarios, to afford these purchasing investments, financial resources in terms of individual budgets $b_i \in B$ are available periodically. As a result, the acquisition process is restricted by this financial constraint that has to be respected always by the total cost.

In summary, the goal of an individual static acquisition problem is to select an optimal plan $\hat{p}$, while at the same time respecting the budget constraint on the total cost $C_\Psi(H)$. This could formally be expressed as:

$$\hat{p} = \underset{p \in P}{arg\ max}\{\mathcal{F}(p_i), C_\Psi(H) \leqslant B\} \tag{5.2}$$

where *arg max* are the arguments of the maxima. These comprise the points of the domain of $P$ at which the function $\mathcal{F}$ is maximized. Since a plan is comprised by a schedule and a subset of cells, the final goal can be also defined as the finding of a schedule $\psi \in \Psi$ and a subset of green areas $\tau \in T \subseteq V$ that effectively use the budget $b \in B$ in order to maximise a predefined objective function $\mathcal{F}$ which assesses the utility of the corresponding plan. This function $\mathcal{F}$ quantifies how well the pursued goal is accomplished at the end of the prediction horizon $H$.

From the perspective of provision of green services, this function can be aimed at solving a covering problem, also called Maximum Service Distance (MSD) (Toregas et al., 1971), where a set of elements needs to be covered with a minimum number of subsets, subject to some constraints. An element is considered covered if it is located within a specified distance from one of these subsets. In this regard, a given scenario can be configured by the set of facilities of a certain type, specifically: green areas located at a given distance from a central point. This central location is represented in this case by the CBD that attracts most of the dwellers, since typically in an urban scenario population decays with distance. The final goal consists of maximising the number of inhabitants who are located relatively close to this type of service, in this case green areas.

Regarding the level of complexity of this type of problem, it can be shown through a reduction of the Maximum-k Covering Problem (MCP) (Ageev and Sviridenko, 1999), that even for a unique time step of the problem, selecting the set of patches of land that maximise the acquisition probability of the policy is an NP-hard optimisation problem (Krause et al., 2014).

Furthermore, the consideration of a single optimisation step can hardly accomplish the final long-time objectives of such plans, the problem should be formulated instead by a sequential set of the previously defined static problems. The management of the financial resources between time steps can be defined in a way that the unused budget assigned in a given time step $t-1$, denoted by $b_r$, is added over the following period $t$ to the corresponding budget $b_0$ as:

$$b(t) = b_0(t) + b_r(t-1) \tag{5.3}$$

where $b(t)$ is the absolute amount of resources available for time step $t$, $b_0(t)$ is the new financial quantity added to the same time step $t$ and finally $b_r(t-1)$ are the remain resources that were not spent in the previous time step $t-1$ and accumulate for $t$.

Apart from that, since land acquisition costs may change over time and additionally urban dynamics could transform areas in the fringe of the city into new developments, which made them inappropriate to be included within any acquisition plan, different patches of land can be available in each time step $t$. Consequently this problem cannot be solved statically in advance without basing the new decisions on previous actions and the forecast of future tendencies.

Under these circumstances, let $P_t \subseteq V$ be the set of patches of land that is available to be purchased in a given time step $t$. For each new time step $t$, the amount of available resources both new $b_0(t)$ and old $b_r(t-1)$, the cost dynamics and the amount of non-urbanised areas refine the set $P$ for time $t$, taking also into account the areas already selected, in such a way that $P_t = T_{t'-1} \cap V$.

## 5.3 Greedy Techniques

### 5.3.1 Adaptive Submodularity

One way of characterising different optimisation techniques is by the level of complexity of the problems that they can deal with. This level is directly linked with the performance of the algorithms and the amount of computational resources required to tackle them.

Previous studies focused on solving a simple static MCP problem concluded that when the problem is solved by simple greedy techniques, most of the time these techniques are able to find the optimum for a two element set (Cornuejols et al., 1980) and that approximate greedy algorithms can solve it within a factor of $1 - (1 - \frac{1}{p})^p$ (Cornuejols et al., 1977) where $p$ is the maximum size of the subsets.

However, when instead of aiming at a simple static MCP problem, the planning of a set of non-static MCP problems is considered, the complexity of the problem, which could be defined as a SDMP, considerably increases and previous conclusions must be revisited. Golovin et al. (2011) conclude that, under some natural conditions, a policy generated by the application of a simple greedy algorithm that searches for an opportunistic location of a facility subject to a budget which is given in the current moment, achieves a performance that can compete with the use of other more intelligent heuristics. These heuristics take into account future trends in the environment like the availability of the land.

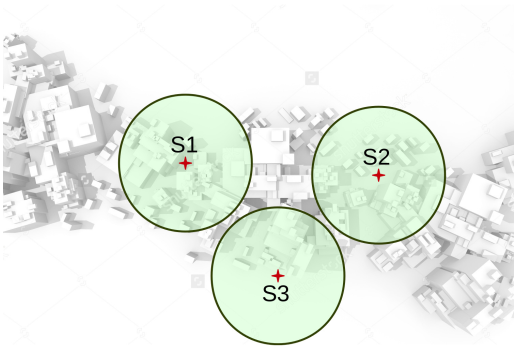
These conditions are met when the function is *adaptive submodular*. Submodularity is a property which assigns each subset $S \subseteq V$ a value $\mathcal{F}(S)$. $V$ in this case is a finite set of the possible locations for a green area and $\mathcal{F}(S)$ is the function which retrieves the utility achieved when the location $S$ is transformed into a park. Assuming that $\mathcal{F}(\emptyset) = 0$, which means that no value is generated in areas without a park, these kinds of functions have a natural diminishing returns property, so that for each subset $A, B \subseteq V$ the following inequality is held:

$$\mathcal{F}(A \cap B) + \mathcal{F}(A \cup B) \leq \mathcal{F}(A) + \mathcal{F}(B) \tag{5.4}$$

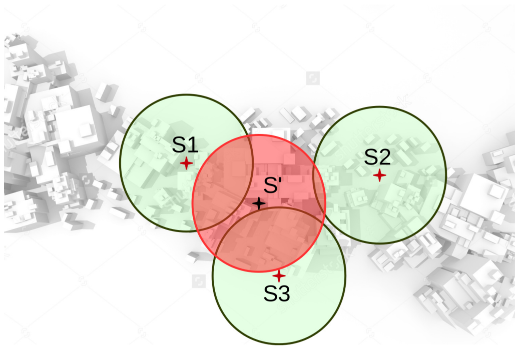
From the previous formula it can be inferred that when spatial overlapping occurs among the areas covered by A and B, there is less gain in the final utility function $\mathcal{F}$. Fig. 5.1 illustrates this effect. A type of submodular function is considered monotone if an augmentation of the number of subsets, does not provoke a drop of the utility function. This can be formally expressed as:

$$\forall A \subseteq B \subseteq V, \qquad \mathcal{F}(A) \leq \mathcal{F}(B) \tag{5.5}$$

For this specific covering problem it is not possible that the transformation of a parcel of land into a new park could provoke a drop in the level of provision of green services in the surrounding area, independently of the number of parks that already exist in the area. Due to that assumption, such location problems are considered to be monotone submodular (Frieze, 1974).




(a) Illustration of the areas of influence of three parks $(S_1, S_2, S_3)$ located in the same city. These areas are strictly disjoint. Consequently no overlap zones are formed from their union. The final utility function will be the result of adding each of the individual utilities separately.




(b) In this case a new green area $S'$ is located among the previous parks $S_1, S_2$ and $S_3$. This concrete placement of $S'$ will include several areas in which the influence of $S'$ partially includes others covered from previous selections. These common areas, as they already serve the inhabitants living within them, do not add new utility to the policy, limiting the contribution of the new green area $S'$. This is an example of a diminishing return effect.

Figure 5.1: Visual representation of the diminishing return function that is linked with the concept of submodularity. In the example it is shown that posterior selections cannot achieve better utility but the same or lower in certain cases.

Submodular functions characterise a broad range of applications including weighted coverage functions, mutual information and facility location among others. Numerous works, aimed at minimising these functions, exist as a result of the amount of areas where this type of function can be applied (Fujishige, 2005). For instance, specialised Branch & Bound algorithms with limited scalability were suggested (Nemhauser and Wolsey, 1981; Goldengorin et al., 1999). However, all of them are limited by the complexity of the problem which, in most cases, belongs to the NP-hard category. This is the case for the weighted coverage (Feige, 1998) and mutual information (Krause and Guestrin, 2012).

Despite this level of hardness, (Nemhauser and Wolsey, 1978) were able to design an intelligent greedy algorithm which achieves good approximations to optimal solutions for several classes of submodular functions. Their approach is not only capable of dealing with a NP-hard complexity, but also competing at the same level as the best results provided by other much more intelligent approaches.

From this point onwards, the concept of adaptive submodularity of Golovin is a generalisation of the classical notion of submodularity for sequential decision making problems under uncertainty where the system is capable of taking into account information from the environment after selecting each element.

Apart from the complex nature of this kind of problem, uncertainty entails that the state of the system is only partially revealed. Due to that and the common presence of varied types of constraints, further difficulty for the optimisation process is added. However, according to Golovin, this structural property ensures robust performance for greedy approximation approaches when dealing with these SDMPs.

For simple problems, greedy algorithms are very efficient optimisation strategies compared with other more intelligent approaches, due to their limited amount of resources required to apply them. However, in order to solve NP-hard problems these techniques generally require exponential time and subsequently they generate very poor results. Golovin demonstrates that, in the cases in which two conditions are met, a simple brute force strategy can be competitive with the optimal policy. These two conditions are adaptive monotonicity and adaptive submodularity. As previously mentioned, monotonicity is the non-negative nature of the benefit of any action and submodularity is based on the idea that the benefit of any action can never be higher if the action is chosen later in time due to its diminishing returns property. The law of diminishing returns in economics says that, if all productive processes are held constant except one, adding more of this factor will at some point yield to lower marginal output of the production process (Knight, 1944). In the same way, let a policy be composed of a series of locations selected in a period of time, if all the selected locations are kept constant except one, which will be added later on to the policy, the final result of this modified policy will never be better than the original one. This is a consequence of a problem that can be characterised as adaptive submodular.

Since in the previous definition of the problem both conditions are met, then an intelligent greedy algorithm that searches for maximising the current utility can compete with other advanced approaches like our evolutionary algorithm. This happens not only in terms of objective values achieved, but also in terms of the efficiency of the entire process. In this chapter, it is shown different configurations of the urban problem when this condition is possible to be met.

### 5.3.2 Model Complexity

The problem could also be analysed from a more abstract point of view, focusing on the inner complexity of the planning problem. This complexity can be derived from different factors. One of these sources is the representation ability of the selected urban model. In this case the model is used to develop the different dynamics that interact within it including population distribution, urban spatial growing patterns and different land-use prices. From this perspective Wheaton (1979) and Berry and Kim (1993) stated that the simplicity of Alonso's model directly affects the nature of the search space and may condition the optimisation process.

EA techniques, compared to other methods, have been shown to be especially suitable for large and complex (nonconvex and nonlinear) search spaces with a very large number of parameters (Aerts and Heuvelink, 2002) and, within a simple environment configuration, EA may not be able to use all its potential. Subsequently other less sophisticated optimisation methods can outperform it in terms of performance since EA achieves better results as the complexity of the planning problem grows. This was one of the conclusions drawn by Pukkala and Kurttila (2005) from comparing six different heuristics for forest planning problems when spatial objectives are included in the model. Spatial objectives require that the optimisation process has to deal with physical locations of the facilities to allocate them. Pukkala and Kurttila (2005) also reported the slow nature of the EA strategy. In this regard, Palahí et al. (2004) added that EA is not good for simple location problems.

From this point onwards, this chapter will be focused on studying the comparison between two adaptive greedy optimisation strategies and an Evolutionary algorithm, using an urban scenario. In response to the offline nature of the EA approach, the algorithm has to deal with this problem assuming significant levels of noise and uncertainty, a complicated objective function and constraints. It is important to notice that in this chapter, the validity of the model to deal with real-world scenarios is not the most important element taken into account for the analysis of the algorithms. Finally, the results show how increasing the level of complexity of the spatial structure of the urban model and the elements interacting within it, the EA approach can compete and outperform the most intelligent greedy algorithm.

## 5.4 Methodology

The methodology followed in this chapter is visually depicted in Fig. 5.2. Green squares represent the different optimisation strategies, orange items are external parts which are computed separately and finally blue squares depict different steps of the planning process according to the varied nature of each strategy (adaptive or non adaptive). The elements that are surrounded by a rounded rectangle represent steps of the study where the hypothetical urban model is involved. After the set-up phase is finished and the statistical data is generated, the workflow is divided into two main areas: on the right, two adaptive baselines used for comparison purposes and, on the left, it is located the non-adaptive evolutionary approach. The last confluence box will gather and compare the final set of policies, resulting from each optimisation process.

Each of the components represented in this workflow are explained in the following subsections.

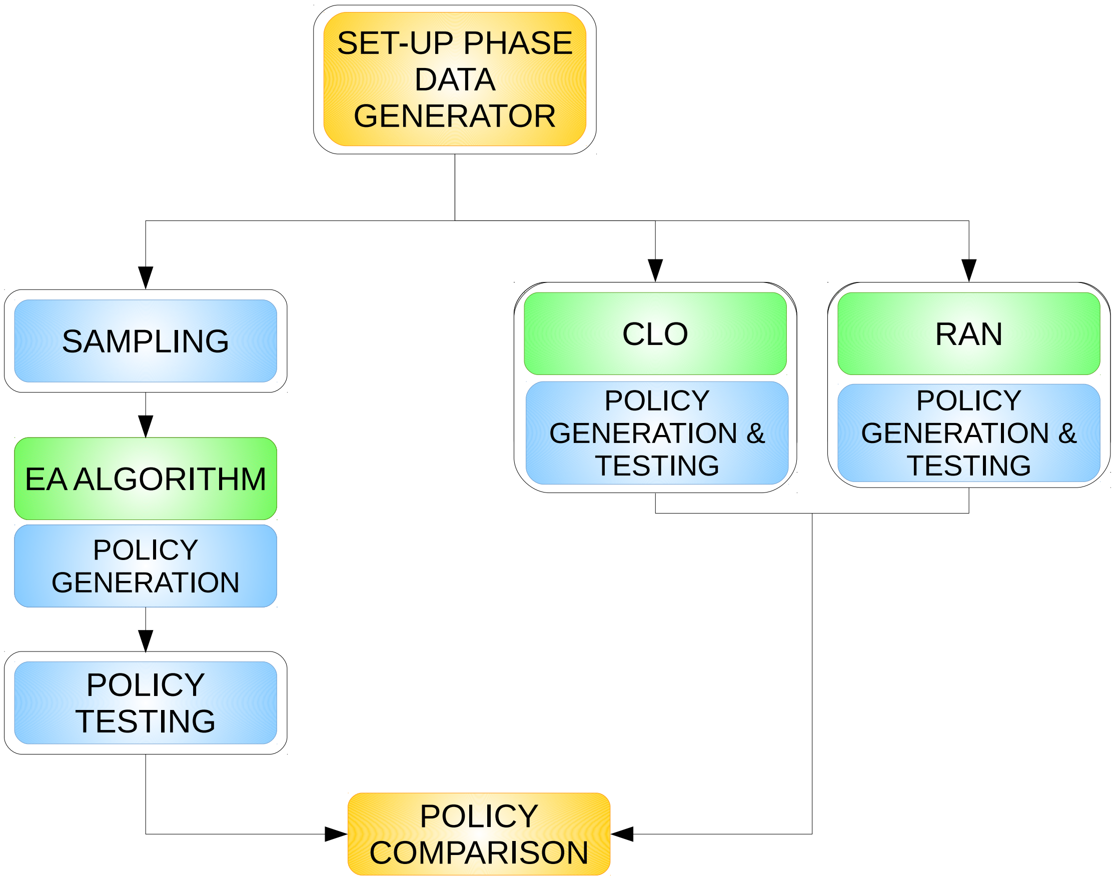



Figure 5.2: General schema of the components used in this work and how they interact with each other.

### 5.4.1 Set-up Phase & Sampling

The workflow that connects the multiple components which encompass the present optimisation process start with a set-up step where certain variables are initialised. In this phase, the common characteristics of all the modules included in the system are defined in such a way that a posterior measurement of the final performance of the optimisation strategies can be carried out. These are, for each experimental run, two initial constraints are imposed on all the approaches considered:

- The budget: A stochastic series of monetary incomes, generated in advance by a uniform random process, is assigned to the *municipality* in each sequence of time in order to implement and manage suitable open space policies. The municipality is the entity in charge of taking the purchasing decisions. The principle behind this conception is that since financial resources assigned to green protection generally are available in increments over time, it is unrealistic to be able to purchase a large set of parcels at once.

- The ecological scenario: The initial ecological configuration of the lattice is defined by a random generation process where an individual value is assigned to each cell of the lattice. These values represent the ecological richness of the area under consideration. An illustration of a final ecological configuration is shown in Figure 3.14.

Both factors affect the land purchase dynamics, the first one limiting the location and amount of parcels of land that the system is able to acquire in each time step, and the second one influencing rural land types. In this regard, the ecological value of a given patch of land defines its configuration as an agricultural or forest land type. Furthermore, the nature of the prices for each of these types is significantly different.

Finally the statistical data gathered for sampling are also generated, as was explained in the previous chapter, see Sec. 4.5.1 on page 135. The information includes which cells are urbanised and when, prices of the urban cells for the entire duration of the simulation and the population dynamics, which include the density of each urban cell for the same period of time. These different sets of data are used to aid the EA algorithm to deal with the noise and uncertainty resulting from its offline nature and also to calculate the fitness of each solution.

### 5.4.2 Evolutionary Algorithm Definition

In this work, an EA approach is used to find a time series of cells to be transformed into parks during a concrete period of time. The optimisation process is highly constrained by geographical and economical circumstances and it should be adapted to work with high levels of uncertainty derived from the unknown future.

To deal with this uncertainty, the system retrieves offline statistical data collected by the same urban simulation which is used as a surrogate model (Vallejo et al., 2013). These different sets of data, gathered from the beginning to the end of the period considered in the planning process include information about: (1) parcel availability, (2) non-urban land prices and (3) population distribution dynamics. Green parcel availability is restricted by the non-deterministic urban dynamics of the city, the evolution of the non-urban prices and the quantity and distribution of the population which is ruled by a utility function that depends on other factors of the model. A formal description of the utility function is depicted in formula 3.4. These elements are used to measure the final satisfaction of the population.

The algorithm starts with the random generation of an initial population of potential solutions, each referred to as a 'chromosome'. Each individual solution stores the information needed for encoding a complete problem solution. The initial population has to homogeneously cover the search space of solutions and ensure enough diversity to successfully allow its future exploration. A lack of diversity at this stage may cause stagnation and premature convergence (Eshelman and Schaffer, 1991; Goldberg, 1989). However, in the present case, a relatively small population size was selected, equal to 25, since by empirical experiments we concluded that in our framework the evolutionary algorithm can guarantee most of the times the reasonably fast location of a near-optimal solution. Larger configurations only elongate the optimisation process without providing noticeably better performance.

The structure of the initial population of solutions and its evolution have always to satisfy some defined constraints in terms of budget and land availability. Green protection is not allowed in urban areas.

#### 5.4.2.1 Chromosome Encoding

Every individual solution, also called a chromosome $c_r \in C_r$ is composed by 'genes', each one denoted by $\beta$ and referred to in this work as an individual selection. They describe the selected cells to be transformed from rural areas into protected units at a certain time $t$. The chromosome structure is visually depicted in Fig. 5.3 and can be formally expressed as follows:

$$c_r = [\beta_0, \beta_1, ..., \beta_{T-1}] \tag{5.6}$$

where $T = 600$ is the number of time slots defined in the simulation, in which the algorithm has the opportunity to perform a purchase if it is considered appropriate. Due to the theoretical nature of the model, an individual time unit in the simulation is not equivalent to a real period of time. A given $\beta_t$ represents the selection of rural cells to be protected at a certain time $t$ (see Figure 5.3). This, in turn can be defined as follows:

$$\beta_t = \left\{[c_{(x_1,y_1)}, c_{(x_2,y_2)}, ..., c_{(x_m,y_m)}], b_t\right\} \tag{5.7}$$

where $c \subset C$ is the subset of $[0, m]$ cells selected for protection by the *municipality* agent at time $t$. Each cell $c_i$ is spatially located in a different pair of coordinates $(x, y)$ to be transformed into a park in turn $t$. $m$, which denotes the amount of cells protected in this turn, is bounded by the maximum *budget* available for this time step $t$. The system will try to buy cells as long as the budget is high enough to allow new acquisitions. In the equation, $b_t$ denotes the remaining budget that was not spent in time $t$ and it accumulates for the next selection $\beta_{t+1}$. The budget $bd \in B$ is generated in advance by a uniform random process taking values in fixed intervals and it is shared by all the optimisation strategies. In each time step $b_t$ can be calculated as:

$$b_t = bd_t - \sum_{i=0}^{m} price(c_i)_t + b_{t-1} \tag{5.8}$$

where *price* is the function that calculates how much a single non-urbanised cell $c_i$ costs according to the formula 3.17 for time $t$, $m$ is the number of cells protected in this time step and $b_{t-1}$ is the remaining budget from the previous period.

This chromosome representation is order-dependent, with the total number of possible facilities to be protected not fixed in advance. Because of some of the selections could be empty, this implies that chromosomes could also be considered of variable length. However, for simplification purposes they are treated as regular.

The current encoding system allows the existence of redundant individuals within the population, which could potentially lead to a lack of diversity during the evolutionary process. Additionally, the encoding process is strongly dependant on the set of selected slots of time where purchases are carried out. These time steps are chosen in the initialisation of each individual solution at the beginning of the optimisation by a random process subject to the budget constraint. The mutation operator will only search for better acquisition areas at these times and no exploration of purchasing opportunities during a different subset of time-steps is done within the EA algorithm.

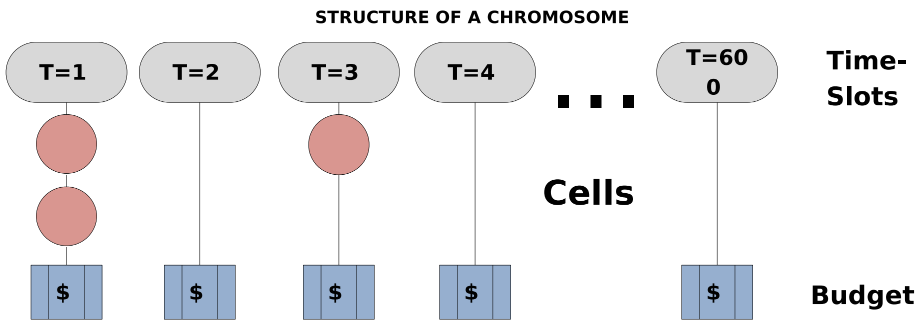



Figure 5.3: Visual representation of the three-layer chromosome encoding used in this approach. The first layer coloured in grey represents a given time-slot defined by its intrinsic order. Afterwards, the set of cells planned to be acquired are depicted in pink circles. Finally the available spare budget that remains after the purchases are done within this time step is shown in blue squares.

Under the current encoding system, the generation of solutions and their evolution always ensures the feasibility of the entire population during policy construction. However, when this policy is tested afterwards, it is possible that unforeseen circumstances arose. Some of the land selections can be infeasible due to a wrong perception of the budget or the urban dynamics of the city when the policy is executed. In this case, the set of feasible solutions $\mathbb{F}$ is a subset of the possible solutions generated by the algorithm that satisfies the problem constraints.

The testing of the feasibility of the solutions is not done at the level of the individual encoding directly. Instead, a list of forbidden locations is used to avoid protecting a cell twice. Another possible approach for this problem would be to encode the entire lattice (a matrix of $50 \times 50$ positions) for each individual and indicate the moment at which the cells are protected. This method would avoid the necessity of having external data structures to store the additional information and would save the time of checking that a given location is used or not. The drawback of this approach would be the storage required to keep the entire population of solutions in memory during the evolution process.

Redundancy in the population of solutions is a factor not taken into account during the optimisation process. Then, the system allows the evolution in parallel of identical solutions.

#### 5.4.2.2 Scheme Selection

The selection operator emphasises good solutions to the detriment of weaker alternatives by the implementation of an elitism mechanism. Elitism in an EA context, is a mechanism that ensures that the best-so-far individuals survive from generation to generation. For selection purposes it was used Tournament Selection, which has shown to lead to good performance, despite its simplicity, and it is frequently used (Nicklow et al., 2009). Tournament size was set to 4, which means that the method selects a parent from the population by first choosing four chromosomes uniformly at random, and selecting the best of those four, breaking ties randomly.

#### 5.4.2.3 Mutation Process

An elitist mutation operator which modifies a single selection is used to find new points in the search space. The implemented mutation method randomly selects one non-empty selection $\beta_i$ and searches for a set of cells that improves the fitness with respect to the same selection $i$ in the parent. Selecting any of the cells already included in the entire policy is not allowed.

Let $\beta_i$ be the selection $i$ picked to be mutated in time $t$ from the parent $c_r$, then $\beta_t^{'}$ will be the new selection to substitute for $\beta_t$ if the following conditions are met:

$$\begin{gathered} f(\beta^{'})_t > f(\beta)_t \\ \beta^{'} \notin \overset{*}{c}_r \\ \mathcal{P}^R(\beta^{'})_t \leqslant \mathcal{P}^R(\beta)_t \end{gathered} \tag{5.9}$$

where $f$ defines the fitness function which assesses the adequacy of the selected cells included in a single selection $\beta$ at time $t$, $\overset{*}{c}_r$ is the chromosome which represents a copy of the current parent solution $c_r^{\prime}$ and $\mathcal{P}^R$ is the total rural price of the patches of land included in a given selection at time $t$.

These conditions can be interpreted as follows: only if the new selection of cells $\beta'$ is strictly disjoint from the cells already selected within the entire policy, achieving at the same time higher satisfaction without violating the affordability terms, then the change is accepted and hence the mutation is successful.

Due to the plateau nature of the search space, the optimisation algorithm encounters significant difficulty when trying to improve the policy between generations. To overcome this, the algorithm boosts the appearance of successful changes within the evolutionary process by performing 1000 attempts in each generation to find an improved mutation, which can be seen as a system which implements two stages of evolution.

Apart from those conditions, there is another factor that complicates the consequences of the mutation procedure. The time-dependent nature of the budget further constrains the price requirements for the new selection. As previously mentioned, the amount of resources assigned to each slot of time $t$ is made up of the corresponding budget for this time slot plus the remaining resources resulting from previous purchases. Then, since the part of the budget that the selection $\beta_t$ has not spent in time $t$ is accumulated for time $t+1$, it could easily happen that any of the next non-empty selection from $\beta^{t+1}$ to $\beta^T$ uses these extra resources for its own purchase. If the mutated selection $\beta'$ uses more resources than $\beta$, then future infeasibilities can occur when the next acquisitions are carried out, invalidating some of the future purchases when the policy is finally executed. If the number of generations that the algorithm uses to evolve its population is large, this effect can easily spread which can provoke numerous rejections when finally the policy is implemented. Other reported approaches eliminate the consequences of this aggregate behaviour by discarding the budget that was not spent in each time slot (Golovin et al., 2011).

A successful mutation should also meet other constraints:

- Cells cannot be selected twice or more for the same individual solution.
- The remaining budget should be always positive. Debts are not permitted.

- Cells being considered for protection cannot have the state of urbanised or be in the process of changing to that state.

- A gene mutation, over a specific selection, should not modify the rest of the genes.

The existence of infeasibilities at the level of single selections when the policy is designed, on the other hand, could be positive at policy execution time under certain circumstances. The algorithm includes a parameter which may slightly relax these budget constraints in the mutation process. It was empirically proven that this mechanism may boost the final performance of the algorithm by incrementing the final number of protected cells and compensating for the ones that are finally rejected by failing to foresee the urbanisation dynamics. In the execution of the policy, a cell can be rejected due to two factors: a lack of budget or when the objective cell has been already urbanised. The algorithm includes a parameter which may slightly relax the budget constraints in the mutation process, introducing a reduced number of infeasibilities. In the case of an urbanisation failure, the resources assigned to the purchase will not be used and can be invested in some extra cells selected by this mutation mechanism.

The pseudocode of the implementation of the mutator operator is depicted in Algorithm 5.1.

In the algorithm, *tick* represents the random time step selected to be mutated. *newBudget* is a local variable used to calculate the funds available to be spent in the cells selected by the new mutate individual. *remainBudget* and *budget* are the funds coming from previous and current time step respectively. *sel* represents the new mutated selection candidate to be included in the policy. *newFitness* is the fitness of the new selection *sel* and *formerFitness* the fitness of the parent. *changed* is a boolean variable triggered when a successful selection is found, *newPrice* is the total price of the cells chosen by *sel*. Finally, *counter* is a variable that sums up the amount of unsuccessful mutations and *MAX_ATTEMPS* the maximum number of attempts that the system allows, which is equal to 1000.

To study the complexity of the mutation operator requires also to analyse other

**Algorithm 5.1** EA Mutation

```
Require: tick, formerFitness, remainBudget, budget
global variables
    MAX_ATTEMPS
end global variables
local variables
newBudget
changed
newPrice
counter
newFitness
sel
end local variables
procedure MUTATE
    newBudget = 0;
    do                                        ▷ Select randomly a tick to mutate
        tick = this.selections.get(new Random()).getTick();
        if tick!=0 then
            newBudget = remainBudget[tick -1];
        else
            newBudget += budget[tick] - remainBudget[tick];
        end if
        sel = new Selection(tick, newBudget); ▷ if there are cells in the selection
        if sel.getCells().size()!=0 then
            newFitness = sel.calculateFitness();
            double formerFitness = this.selections.get(index).getFitness();
            if newFitness > formerFitness then
                changed = true;
                newPrice = 0;
                for Cell c: sel.getCells() do
                    newPrice += c.getPrice();
                end for        ▷ Delete previous selections from the forbidden list
                for Cell c: this.selections.get(tick).getCells() do
                    deleteForbidden(c.getX(), c.getY());
                end for
                selections.set(index, sel);
                remainBudget[tick] = newBudget - newPrice;
            else
                counter++;
            end if
        else
            counter++;
        end if
    while !changed && counter < MAX_ATTEMPS
    if counter < MAX_ATTEMPS then
        return true;
    else
        return false;
    end if
end procedure
```

methods and functions that this code uses in each execution call. There are two functions that require further analysis: selection constructor has a loop that can be repeated a maximum of 100 times and the function *createCell()* another 40 times. Then, according to this information, in the best case the time complexity of the algorithm is $\mathcal{O}(1)$ and in the worst case $\mathcal{O}(n^m)$, where $n = 1000$ and $m = 140$.

#### 5.4.2.4 Crossover Operator

In the present implementation, the selected structural representation of individuals introduces difficulties for crossover operations. Firstly, as a rural cell cannot be protected twice, a feasible crossover needs to ensure that the sets formed by the cells included into the segments to be integrated in the new chromosome are strictly disjoint.

This restriction can be manageable to implement. However the time-dependent nature of the remaining budget also highly constrains the system, as occurred in the implementation of the mutation operator. To minimise the number of possible infeasible selections, the following condition should be met. If a pair of parental chromosomes $C_r$ and $C'_r$ are selected to mate and recombine in such a way that $\beta_i$ is the binding crossover point and $B_i$ the remaining budget for $C_r$, and in the same way, $\beta_j$ and $B_j$ are the corresponding parameters for $C'_r$, then $B_i > B_j$.

The high likelihood of joining two incompatible chromosomes leads to difficulties in easily finding a suitable combination to compose the new offspring. In this regard, this search of compatible individuals causes an exponential increment in the computational time of the algorithm. As a conclusion, the crossover operator was removed from the system, even if this decision could also reduce the time performance of the algorithm (Jansen and Wegener, 2002).

#### 5.4.2.5 Fitness Function

In the current chapter a single objective is considered. In the model, the fitness is defined to reflect how close agents' homes are to green spaces, and involves measuring the distance from the dwelling of each agent to the closest green area located in the surroundings. This metric is based on the premise that distance is the factor that mainly determines the access and the frequency of use of green areas (Giles-Corti et al., 2005; Lindhagen, 1996; Woolley, 2003). According to the specific criterion of Giles-Corti et al. (2005), green areas can be classified into the following groups:

- Access within a short walk (less than 300 metres).
- Access within a long walk (from 300 to 600 metres).
- Access with help of any means of transport (further than 600 metres).

The same study concludes that people do not generally use a green area if it is located beyond a threshold of 300-400 metres from their residence. Using this convention, the optimal candidate solution is the chromosome which maximises the usability of green areas in compliance with the multiple constraints of the system such as the restricted budget and land availability. Notice that even if the problem under consideration is considered a *covering* problem, this concept is not included specifically in the calculation of the fitness. However, since the definition rewards the areas closest to the dwelling of the agents, the covering idea is implemented indirectly.

According to this idea, fitness is formulated as follows: let $A$ be the set of agents living in the city and $G$ the set of green areas located in the urban area under consideration, then the fitness value $f(a_i, g_j)$ at a time $t$ measures the satisfaction achieved by this concrete agent $a_i$ in relation to the green area $g_j$. If $\forall a \in A$ and $\forall g \in G$, this value can be formulated as follows:

$$g_j = clo(a_i)$$

$$f_t(a_i, g_j) = \begin{cases} 3 & \text{if } \delta(a_i, g_j) = 1 \\ 2 & \text{if } \delta(a_i, g_j) = 2 \\ 1 & \text{if } \delta(a_i, g_j) = 3 \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

where *clo* is the function that returns the closest green area $g_j$ of a given agent $a_i$ and $\delta(a_i, g_j)$ is the function that calculates the distance from a given agent $a_i$ to $g_j$ in the grid using the Manhattan distance metric (Krause, 2012).

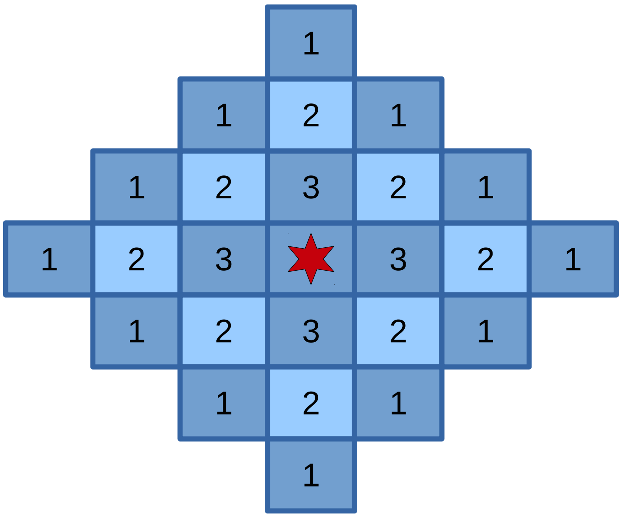



Figure 5.4: Visual representation of the type of neighbourhood generated by the fitness function. Squares represent cells a short distance from a given green area depicted by a red star. Inside each square is included a quantification of the contribution of this urban park to the final fitness in the case an agent settles down in this cell.

Manhattan distance can be computed using the formula $z = (x_i - x_j) + (y_i - y_j)$ and is a metric more suitable for grid-like road networks that tries to mimic restrictive movements typical of the rectangular pattern of streets (Bow et al., 2004; Shahid et al., 2009). A visual representation of the neighbourhood generated by the definition of this fitness function is depicted in figure 5.4. It is important to notice that the coverage provided by a green area falls off gradually following the behaviour of a decay function.

From an urban planning perspective, to appropriately assess the suitability of a planning policy, it is not enough to know how effective the plan is for a given moment in time. Instead it is necessary to determine its adequacy for the entire interval covered by the planning procedure. To accomplish this goal, the fitness of each parcel of land included in each selection $\beta_i$ is measured at a time $t^*$, when it is acquired, to the end of the simulation $H$. Formally this can be represented as follows:

$$F(g_j) = \sum_{t=t^*}^{H} \sum_{i=1}^{n} \big(f_t(a_i, g_j)\big) \tag{5.11}$$

where $n$ is the total number of inhabitants in the city at time $t$ and $F(g_j)$ is the fitness of a single green area $g_j$ calculated for the entire population of the city for the period $[t^*...H]$. Finally if the fitness is collected for the entire set of protected areas at the end of the planning period $H$, then:

$$\mathcal{F}(H) = \sum_{k=1}^{m} F(g_k) \tag{5.12}$$

where $\mathcal{F}(H)$ is the total aggregated fitness of every green area network and $m$ is the number of green areas planned to be purchased within the policy.

Notice that even if overlapping is not directly penalised in the definition of the fitness function, since each agent only contributes to the common satisfaction with the closest green area, the possible rest of parks in the surroundings do not add any value to the function and, hence overlapping of green areas is indirectly penalised.

The pseudocode of the implementation of the fitness function is described the following algorithm, see Alg. 5.2.

**Algorithm 5.2** Calculation of the Fitness Function (Total Satisfaction)

```
Require: minGreenDistance, totalSatisfaction
int minGreenDistance = Integer.MAX_VALUE;
int totalSatisfaction = 0;
procedure FITNESSFUNCTION
    for each agent a in the grid do
        minGreenDistance = lattice.distanceTo(a.location, CellState.PROTECTED);
        if (minGreenDistance > 3 || minGreenDistance==0) then
            greenSatisfaction = 0;
        else
            greenSatisfaction = 4 - minGreenDistance;
        end if
        totalSatisfaction += greenSatisfaction();
    end for
    return totalSatisfaction;
end procedure
```

In the code, *minGreenDistance* is the minimum distance calculated from the location of each agent $a$ to the closest cell protected and *totalSatisfaction* accumulates the satisfaction achieved for the entire population of the grid. The complexity of the function is $\mathcal{O}(n)$, where $n$ is the number of agents in the grid in this time step.

#### 5.4.2.6 Replacement method & Stopping criterion

A steady state replacement method is used to generate the new population in each generation, where fitter offspring are inserted into the current population, replacing less profitable parents. The selected stopping criterion is twofold: the algorithm stops when a maximum number of generations (10,000) is reached or if a certain number of generations (200) have passed without improvement. The latter value was selected after observing, in preliminary work, that once this threshold is reached it is unlikely that further evolution occurs. When the algorithm finishes, the individual solution with the highest fitness is accepted as the optimum result. The offline approach always allows the system to run until convergence is reached.

Finally, when the algorithm finishes, the individual solution with the highest fitness is accepted as the optimum result. At this point this final policy is ready to be tested. The application of this kind of policy is represented by Fig. 5.5.

### 5.4.3 Modules work-flow: Baselines

Two adaptive baselines were used to compare the performance of our non-adaptive evolutionary approach. The first one is called random (RAN), and the second one is the 'closest to the CBD' (CLO) baseline. Notice that comparing adaptive approaches which are free from the influence of uncertainty and are always fully aware of the state of the system may be unfair for the non-adaptive strategy. In this regard, when the offline final plan is applied to the real problem, a reduced percentage of its candidate cells resulting from the application of the EA approach may easily be rejected because the area to protect is already urbanised or because the final price of the cell is higher than expected and unaffordable for the current budget. This testing phase is not necessary for both of the online baselines.

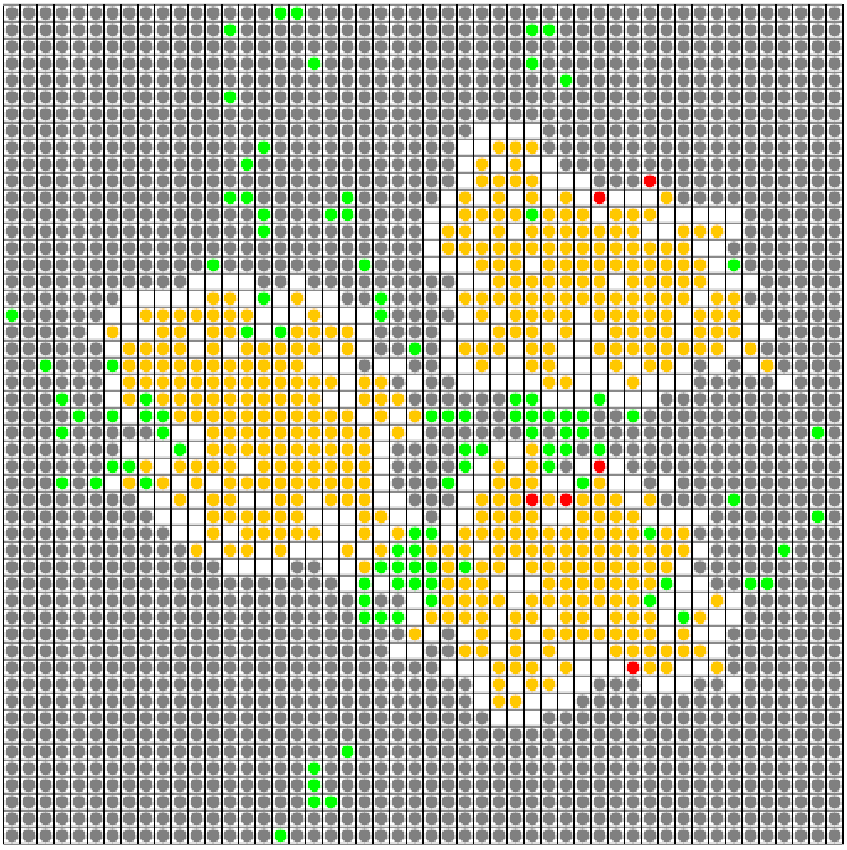


Figure 5.5: In this graphical representation of the application of the EA strategy to the allocation of green spaces, a spatial layout of a city with three CBDs captured at $t = 300$ is shown. The urban cores are depicted in yellow (urbanised areas), white (new areas already transformed to be urbanised but not constructed yet) and red cells (which show the new areas selected in this turn to be urbanised). Green areas are modelled in green. Finally the rest of the grey cells are rural areas which are available for purchasing.

A more detailed description of both baselines is depicted in the following sections:

#### 5.4.3.1 RAN Definition

The first baseline, called *RAN*, is a simple random approach which searches up to ten times in each turn $t$ for rural cells to be transformed into green areas. These cells should hold the following condition: if $R$ is the set of rural cells in the grid at a time $t$, then the system searches for a cell $r \in R$, such that:

$$\mathcal{P}_t^R(r) < b_t \tag{5.13}$$

where $\mathcal{P}_t^R(r)$ is the price of the rural cell $r$ and $b_t$ is the available budget at time $t$. If in all the ten attempts the searches are unfruitful, no purchase is done at this time step and $b_t$ is accumulated for turn $t+1$. A visual representation of the application of this simple policy is depicted in Fig. 5.6.

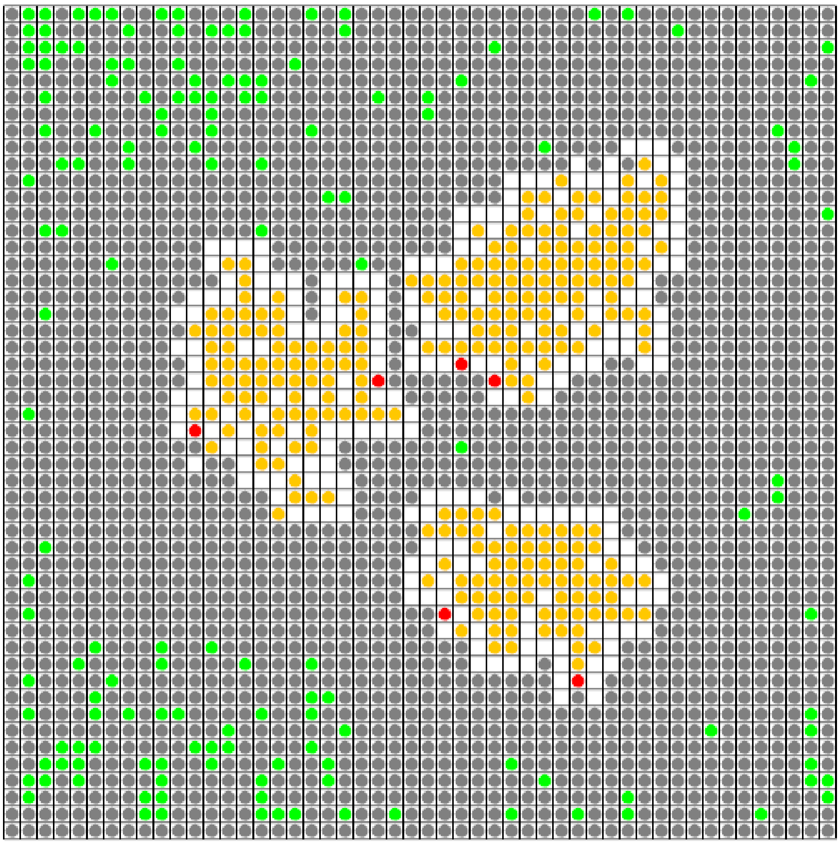


Figure 5.6: In this graphical representation of the application of the online RAN baseline to the allocation of green spaces, a spatial layout of a city with three CBDs captured at $t = 300$ is shown. The urban cores are depicted in yellow (urbanised areas), white (new areas already transformed to be urbanised but not constructed yet) and red cells (which show the new areas selected in this turn to be urbanised). Green areas are modelled in green. Finally the rest of the grey cells are rural areas which are available for purchasing.

The pseudocode of the implementation of the 'RAN' baseline is described in Algorithm 5.3.

In the code, *currentTick* is the current time step of the system, *budget[currentTick]* represents the funds assigned beforehand for acquisition, *accuBudget[currentTick-1]* are the funds accumulated from the previous time step, both types of funds are summed up and passed to the *searchGreenSpace* function as a parameter. The function returns *accuBudget[currentTick]*, that is the remain budget available for land purchases at this time step that will be used in the next future acquisition. In the function *searchGreenSpace*, *selectedCell* is the cell selected randomly for protection if the conditions are acceptable, *MAX_ATTEMPS* is the maximum number of times that feasible cells are sought. This constant has a value of 10. Finally, *attempts* is a counter of the number of times that the system has searched for a suitable cell so far. The time complexity of the algorithm is $\mathcal{O}(n)$, where $n$ is the number of attempts. Since the number of attempts is bounded to 10, which corresponds to the worst case, the time complexity of the algorithm can be approximated to $\mathcal{O}(1)$.

**Algorithm 5.3** 'RAN' Selection of Green Areas

```
global variables
    MAX_ATTEMPS
    SIZE_LATTICE
end global variables
require currentTick, totalBudget, accuBudget[], budget[]
procedure selectGreenAreas(int currentTick)
    if currentTick==0 then
        accuBudget[currentTick] = searchGreenSpace(budget[currentTick]);
    else
        totalBudget = budget[currentTick] + accuBudget[currentTick-1];
        accuBudget[currentTick] = searchGreenSpaceS(totalBudget);
    end if
end procedure
require restBudget, coordX, coordY, selectedCell, attempts
procedure searchGreenSpace(int restBudget)
    int attempts = MAX_ATTEMPS;
    randnum.setSeed(System.currentTimeMillis());
    while attempts > 0 do
        coordX = 1 + (Rand.nextInt(SIZE_LATTICE - 1));
        coordY = 1 + (Rand.nextInt(SIZE_LATTICE - 1));
        selectedCell = GreenArea.getCell(coordX, coordY);
        if selectedCell ! = null then
            if selectedCell.isEmpty() &&
            selectedCell.getPrice() < restBudget then
                restBudget -= selectedCell.getPrice();
                selectedCell.setState(CellState.PROTECTED);
                attempts = MAX_ATTEMPS;
            else
                attempts–;
            end if
        else
            attempts–;
        end if
    end while
    return restBudget;
end procedure
```

#### 5.4.3.2 CLO Definition

The selection of cells in the second baseline called *CLO* is only permitted within the peri-urban area of the city. The philosophy behind this approach comes from the assumption that it is often unnecessary to purchase parcels that are spatially remote from the current population until the city has spread enough to make them relevant. Following this idea, the system retrieves the latest cell urbanised, called $l$. If $N(l)$ is the set of cells that belongs to the neighbourhood of $l$ and $R$ is the set of rural cells in the grid at time $t$, then the system selects randomly a cell $c \in (N \wedge R)$ such that:

$$\mathcal{P}_t^R(c) < b_t \tag{5.14}$$

where $\mathcal{P}_t^R(c)$ and $b_t$ are the price of the cell $c$ and the available budget at time $t$. If none of the neighbours is suitable for protection due to the current constraints, then the budget $b_t$ is accumulated for the next turn $t+1$. Fig. 5.7 represents a possible development of this policy within a city.

The pseudocode of the implementation of the 'CLO' baseline is depicted in the algorithm 5.4:

In the code, *currentTick* is the current time step of the system, *budget[currentTick]* represents the funds assigned beforehand for acquisition, *accuBudget[currentTick-1]* are the funds accumulated from the previous time step. *restBudget* is the remain budget available to search for the next available cell and *isFound* is a boolean value that is triggered if the algorithm is able to find a suitable cell.

In this case, the time complexity of the algorithm is $\mathcal{O}(n)$, where $n$ is the number of cells that belong to the neighbourhood of the cell that was most recently urbanised. Since according to the definition of the neighbourhood structure, the maximum number of cells in a neighbourhood is 8, then the time complexity of the function could be approximated to $\mathcal{O}(1)$.

**Algorithm 5.4** 'CLO' Selection of Green Areas

```
require currentTick, totalBudget, accuBudget[], budget[];
procedure SELECTGREENAREAS(int currentTick)
    if currentTick==0 then
        accuBudget[currentTick] = searchGreenSpace(budget[currentTick]);
    else
        totalBudget = budget[currentTick] + accuBudget[currentTick-1];
        accuBudget[currentTick] = searchGreenSpaceS(totalBudget);
    end if
end procedure
require restBudget, isFound
procedure SEARCHGREENSPACE(int restBudget)
    boolean isFound = false;
    for Cell c in getLastUrbanised().getNeighbours() do
        if !isFound then
            if c.isEmpty() && c.getPrice() < restBudget then
                isFound= true;
                restBudget -= c.getPrice();
                c.setState(CellState.PROTECTED);
            end if
        end if
    end for
    return restBudget;
end procedure
```

## 5.5 Experiments & Computational Results

In this section the implications of the model, when different strategies are applied, are explored analytically. The methods selected for this analysis are an offline EA algorithm and the two baselines presented previously: 'RAN' and 'CLO'. For each of them, the search of an efficient policy constructor is carried out according to its implemented strategy.

The first inspection of the generated data was aimed at exploring the topological arrangement of the protected cells generated for each of the approaches. The different set of land-uses were plotted, see Fig. 5.5 for the EA algorithm, Fig. 5.7 for the 'CLO' strategy and Fig. 5.6 for the 'RAN'. A preliminary visual inspection shows that EA is able to allocate their selected green areas between the different urban cores. The CLO strategy generates more green spaces in the urban area and less at the edges of the grid. The RAN baseline places less green spaces between urban areas and more at the edge of the area under consideration. These differences are mainly caused by the distinct nature of the strategies pursued by each of the algorithms. The RAN algorithm, due to its random nature, tends to finish with a more homogeneous layout of green areas. These patches of land are mostly located in the outer areas of the lattice due to the cheaper prices of the land that is further from the city. CLO, only allows the acquisition of cells at the boundaries of the city, so allows a mix of green areas within the core of the city. Finally EA, is a trade-off strategy between both previous approaches allowing protection at a middle distance from each city centre.

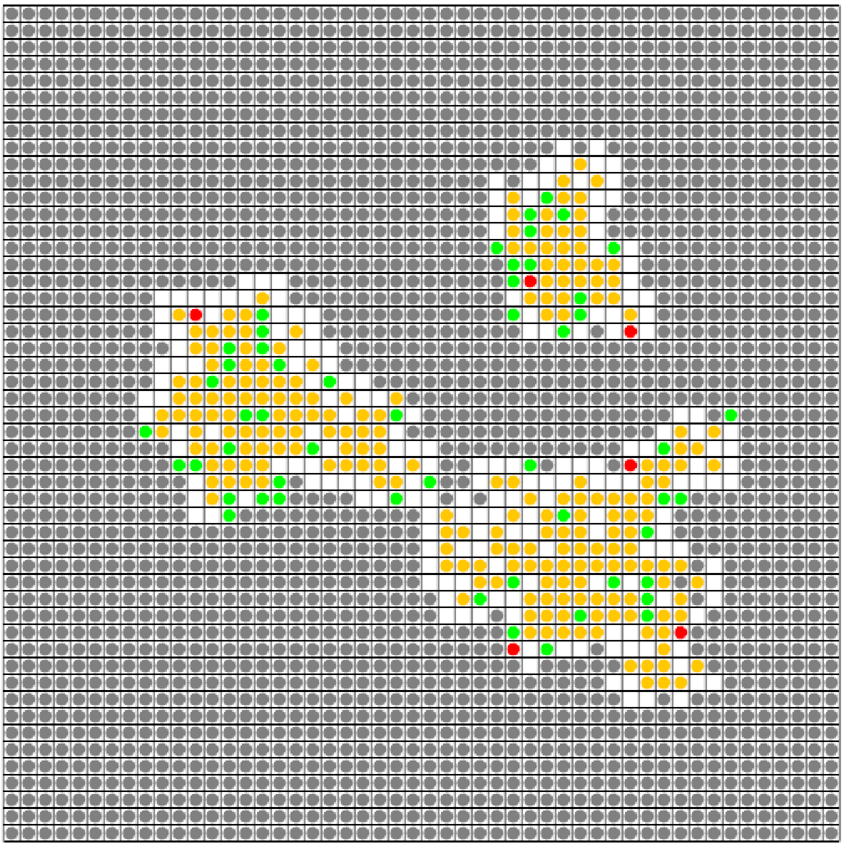


Figure 5.7: In this graphical representation of the application of the online CLO baseline to the allocation of green spaces, a spatial layout of a city with three CBDs captured at $t = 300$ is shown. The urban cores are depicted in yellow (urbanised areas), white (new areas already transformed to be urbanised but not constructed yet) and red cells (which show the new areas selected in this turn to be urbanised). Green areas are modelled in green. Finally the rest of the grey cells are rural areas which are available for purchasing.

Afterwards, the optimisation strategies were tested against a predefined number of scenarios where the urban model was modified to meet different sets of features such as multiple numbers of CBDs growing in parallel and different non-urban price dynamics. The main goal pursued by this strategy is to study the behaviour of the optimisation techniques in relation to the multiple complexities of the urban model, focusing especially on how the EA performance varies in comparison with the baselines. Hence, the scenarios were defined as a measure of the complexity of the selected configurations. Starting from a simple layout and constant prices, increasing the level of complexity of the topological arrangement and finishing adding a dynamic behaviour to the rural prices. Following this increment in the complexity, the characteristics of each scenario are summarised in the following list:

- **Scenario 1** is devoted to the analysis of an urban area with a single and central CBD. Non-urban prices are constant and hence, peri-urban parcels are as expensive as any other rural area in the lattice.
- **Scenario 2** also deals with only one central CBD but, in this case, the complexity is increased by defining dynamic non-urban prices. These prices augment as a function of the growth of the city, based on the reduction of the available land and the corresponding increment in demand and finally in prices.
- **Scenario 3** configures a layout with several CBDs distributed in a triangle shape in the centre of the lattice and with non-urban prices of a dynamic nature.

In order to account for the variations introduced by the stochastic variations, the results presented in this work have been calculated as averaged over 20 independent optimisation runs for each method considered. Different topological configurations will be considered for each of the approaches in line with the scenarios previously described. In each of the runs, different data will be collected to assess and compare the suitability of the strategy. The data include: the satisfaction (performance/fitness of the policy), total population, total urban and rural cells, prices of urban and rural cells (maximum and minimum values), average distance from the green cells to the correspondent CBD and incoming migration.

The selected information analysed for measuring the optimisation approaches comprise the final amount of protected cells that the approach manages to buy with the available budget. The satisfaction of the population is measured by means of the fitness function and the average distance from each protected cell to the corresponding CBD.

The visualisation of the fitness function is a method capable of showing the success of the algorithms in finding an efficient set of green areas which satisfies the necessities of the population. The number of green areas is a quantitative metric that indicates the amount of green areas the algorithms are able to protect with the available budget, independently of their proximity to the inhabitants of the city.

Finally, the closeness factor is an important qualitative metric to be measured in order to properly compare different policies. This metric measures the global distance of the parks to the CBDs of a given policy generated by an optimisation strategy in each time step. Hence, for time $t$ this metric denoted by $CL$ is calculated as follows:

$$CL(t) = \frac{1}{m}\sum_{i=0}^{m} \delta\big(g_i, C_{CBD}(g_i)\big) \quad \forall g \in G \tag{5.15}$$

where $m$ is the number of green areas at time $t$, $C_{CBD}(g_i)$ is a function that returns the closest CBD to a given green area $g_i$ and $\delta$ is a function that retrieves the discrete radial distance from this green area $g_i$ to this CBD.

The reason behind its use in this analysis is based on the idea of finding a trade-off between the amount and quality of green areas. If the model acquires almost exclusively cells located in the outer areas of the lattice, which are significantly cheaper, only when the city expands its boundaries at the end of the simulation horizon $H$, these areas are going to be reached. Then, only a small number of residents can get benefit from them, since most of the population will be closer to the CBD and, hence, this may easily result in the generation of a very poor policy.

In this regard Xue et al. (2012) conclude that when the design of long term plans is performed over a population that tends to spread, it is commonly unnecessary to acquire parcels that are at a significant distance from the source at the beginning of the planning period. Under Alonso's assumptions (Alonso, 1964), this distance could be measured by this closeness factor.

The data depicted in the Figures 5.8, 5.9 and 5.10, which correspond to the aforementioned scenarios, are composed by three plots which represent the fitness function as a measure of the population satisfaction achieved in each time step (top), the amount of cells that have been converted into parks (bottom left) and the closeness metric which measures the average distance of all protected cells to the corresponding CBD in each turn (bottom right). These factors will be the major characteristics taken into account to analyse the results presented here.

### 5.5.1 Scenario 1 (1 CBD - constant non-urban prices)

Due to the constant nature of the non-urban prices during the simulation, the optimisation strategies do not get any benefit from purchasing cells in advance further from the CBD. In addition, when these outer areas are urbanised in the long term, they will be significantly less populated than the cells located closer to the centre due to the monocentricity assumption of Alonso (Alonso, 1964).

In this scenario, the three approaches: RAN, CLO and EA can manage to purchase almost the same amount of cells. Only EA is able to buy slightly more cells due to the advanced characteristics implemented in the algorithm. The stochastic option was further explained in Sec. 6.3.2 on page 203 and the existence of infeasible solutions at the level of individual selection was commented on Sec. 5.4.2.3 on page 161. It should be recalled here that all the approaches always share the same budget.

Under these circumstances, the approach that places the maximum number of cells in the most populated areas of the grid, which means the areas closer to the CBD, will have the most satisfied population. In this regard a heuristic like CLO, which has been specially designed to maximise the closeness to the CBD, achieves notably better satisfaction than the other two strategies. The RAN approach cannot compete with the other greedy strategy because it tends to generate a homogeneous distribution of green cells, placing most parts of the parks in not very populated areas. EA, on the other hand, cannot achieve a closeness rate as high as CLO. Furthermore, EA decisions should be made very accurately when selecting areas that will be very close to the borders of the city in order to avoid later rejections due to non-predicted dynamics in the development of the peri-urban areas of the city. These rejections may lead to the situation where very promising areas for locating parks may not be successfully transformed in the long term.

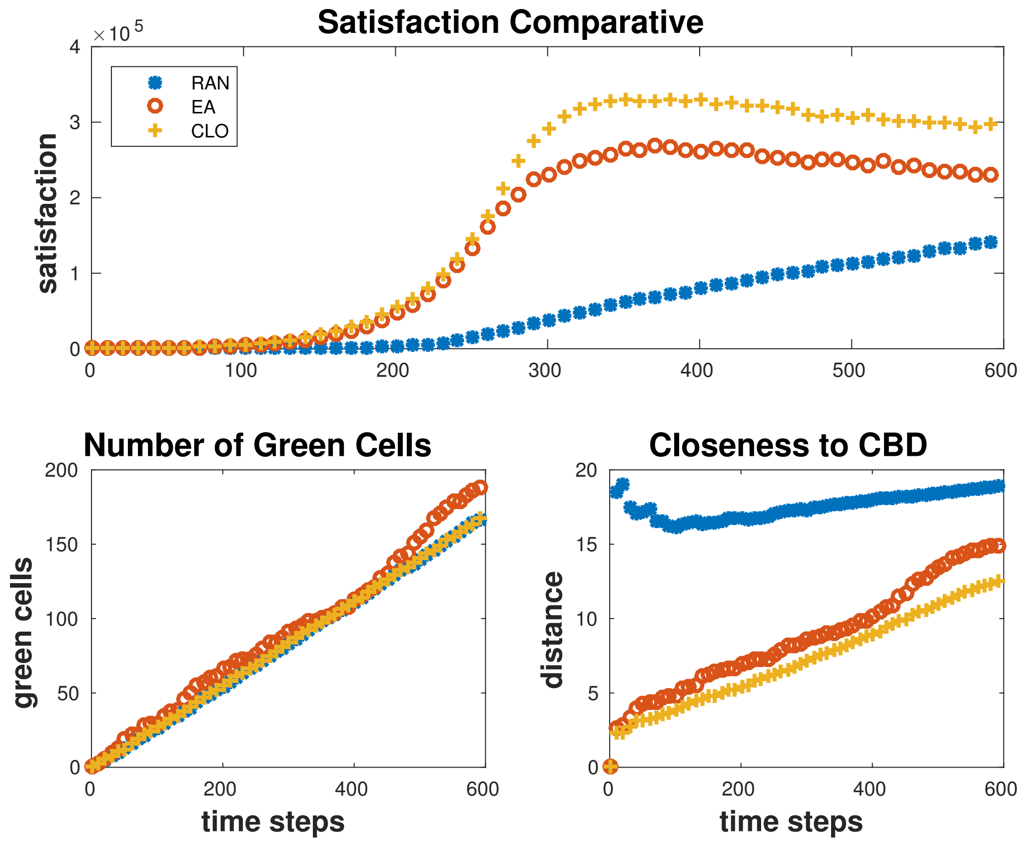



Figure 5.8: Results from scenario 1 (one CBD and constant non-urban prices). The figure shows three different areas where the three approaches are compared: Satisfaction (top), number of cells protected (bottom left) and closeness to the CBD (bottom right). RAN approach is plotted in blue, EA in red and CLO in yellow.

Regarding the number of protected cells, all three approaches manage to buy a constant amount of cells during the planning period, with no clear differences among them. This behaviour is due to the constant nature of the rural cells and the shared financial budget.

### 5.5.2 Scenario 2 (1 CBD - dynamic non-urban prices)

In the second scenario the complexity derived from the dynamical nature of the non-urban prices (See Fig. 3.17) decreases the maximum satisfaction achieved by all the approaches, more significantly in the case of CLO. The noticeable increment in the price of peri-urban areas due to the imminent urbanisation provokes a clear

| TIME STEPS | EA PERFORM. | RAN PERFORM. | CLO PERFORM. |
|---|---|---|---|
| **50** | 271.32 | 3.09 | 428.09 |
| **100** | 2175.44 | 42.95 | 3002.43 |
| **150** | 8419.56 | 346.86 | 10237.07 |
| **200** | 27603.30 | 1669.40 | 32205.26 |
| **250** | 82462.58 | 7575.35 | 88278.98 |
| **300** | 191220.60 | 25374.84 | 218537.87 |
| **350** | 250959.40 | 49536.67 | 315634.73 |
| **400** | 264695.60 | 69811.95 | 328719.40 |
| **450** | 260386.60 | 88600.99 | 322734.00 |
| **500** | 249761.40 | 105596.54 | 312283.87 |
| **550** | 243775.60 | 120505.60 | 302636.80 |
| **600** | 234398.00 | 136816.10 | 296625.87 |

Table 5.1: Numerical values of the EA, RAN and CLO algorithms' performance in terms of the satisfaction achieved by the urban population and measured by the fitness function during the complete time horizon of the simulation (data in line with Fig. 5.8 corresponding to Scenario 1)

distinction in the number of cells that each approach can afford. In this regard the CLO baseline is hindered the most, meanwhile the EA is only partially affected because its strategy tends to compensate for the allocation of its green cells in less dense areas with the acquisition of a quantitatively larger amount of cells. On the other hand, the homogeneous location of the cells allocated by RAN makes the approach almost immune to the increment in prices of the peri-urban areas since the percentage of protected cells in these areas is not very significant.

Since at the beginning of the simulation prices are still affordable for CLO, the method soon manages to acquire enough good cells to gather a significant amount of satisfaction from the population. Meanwhile the results reported by EA are always a step behind. However the situation changes as rural prices increase with time and it is more complicated to acquire parcels with the available budget, specially very expensive peri-urban cells. The value of the available budget can be insufficient in the long term. Since CLO cannot find any suitable cell, its budget cannot be used for any new acquisition, even though its value shows a progressive accumulation in each turn. This effect can be observed at a concrete point in the planning process, close to half of the horizon of the simulation.

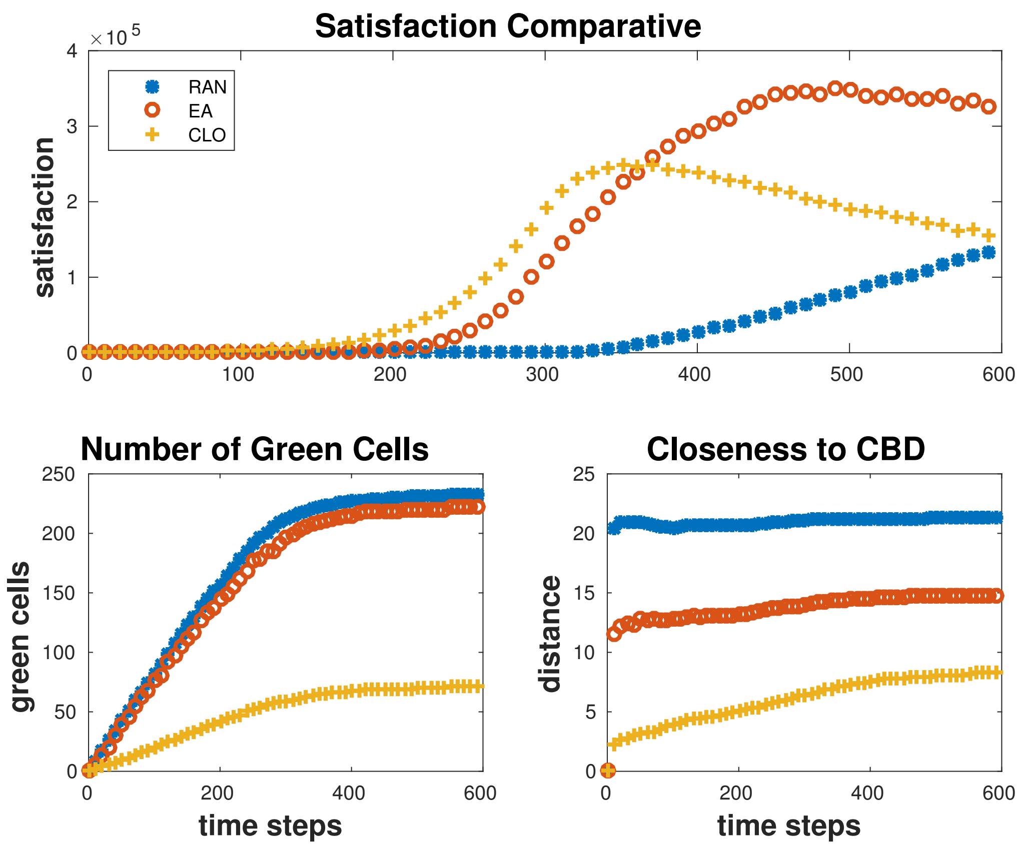



Figure 5.9: Results from scenario 2 (one CBD and dynamic non-urban prices). The figure shows three different areas where the three approaches are compared: Satisfaction (top), number of cells protected (bottom left) and closeness to the CBD (bottom right). RAN approach is plotted in blue, EA in red and CLO in yellow.

There is a peak in the satisfaction from which CLO cannot afford to purchase almost any other new cell on the outskirts of the city, neglecting the provision of parks in these new constructed urban areas. At the end of the simulation, the failure of CLO and a drop in the population who lives close to the CBD provokes a significant decrement in its performance, with its final satisfaction levels equal to the RAN approach. On the contrary, under the same circumstances, EA is at least able to keep its performance steady throughout the period considered.

From the closeness point of view, CLO notably manages to place its green facilities in areas close to the city centre. After that, EA is the second-best strategy, followed finally by RAN. Noticeable is also that, in the three cases, the tendency of the metric is almost steady from the beginning to the end of the time horizon covered by the three planning processes. This means that, even if later in time it is more complicated to find new affordable areas, these new acquisitions are placed in general as conveniently as the ones purchased previously in the plan.

| TIME STEPS | EA PERFORM. | RAN PERFORM. | CLO PERFORM. |
|---|---|---|---|
| **50** | 0.00 | 0.00 | 135.76 |
| **100** | 0.00 | 0.00 | 1335.43 |
| **150** | 93.16 | 0.00 | 4769.86 |
| **200** | 1982.86 | 0.00 | 15976.31 |
| **250** | 13579.84 | 0.00 | 48718.16 |
| **300** | 67806.14 | 4.68 | 125336.04 |
| **350** | 173448.00 | 2898.12 | 226681.00 |
| **400** | 262018.20 | 17655.30 | 246143.50 |
| **450** | 318045.60 | 39203.72 | 228163.90 |
| **500** | 345126.80 | 66414.04 | 204034.80 |
| **550** | 342345.20 | 95586.00 | 182350.80 |
| **600** | 331457.00 | 124703.60 | 162976.80 |

Table 5.2: Numerical values of the EA, RAN and CLO algorithms' performance in terms of the satisfaction achieved by the urban population and measured by the fitness function during the complete time horizon of the simulation (data in line with Fig. 5.9 corresponding to Scenario 2)

For the number of cells finally collected, EA and RAN behave similarly and significantly better than CLO. The difference between them depends on how close the strategies are able to allocate the green cells to the centre of the city which can be seen in the third part of the figure.

### 5.5.3 Scenario 3 (Three CBD - dynamic non-urban prices)

In the last scenario, where more CBDs are included in the framework and the system is more complex, the overall absolute performance of CLO and EA again decreases. RAN is not affected by this new level of complexity since the method tends to spread homogeneously its protected areas independently of the topological structure of the city.

On the other hand, the difference in performance between the CLO baseline and EA is more evident in the long term. The higher complexity of the scenario, with three different urban price gradients, one per each core, creates a more heterogeneous search space where the EA approach can find better solutions, especially at the end of the simulation. The peak point when CLO starts to decrease, is located earlier in time than in the previous scenario, which confirms how the algorithm struggles earlier within a more complex scenario. The opposite behaviour is shown in the EA strategy, where the approach is capable of keeping a growth tendency for longer. These peak points give information about when the algorithm starts having problems to find a proper cell to protect. Under the same conditions, in terms of rural prices, if in one scenario this occurs early in time, this is a signal that the level of complexity is higher than in the other.

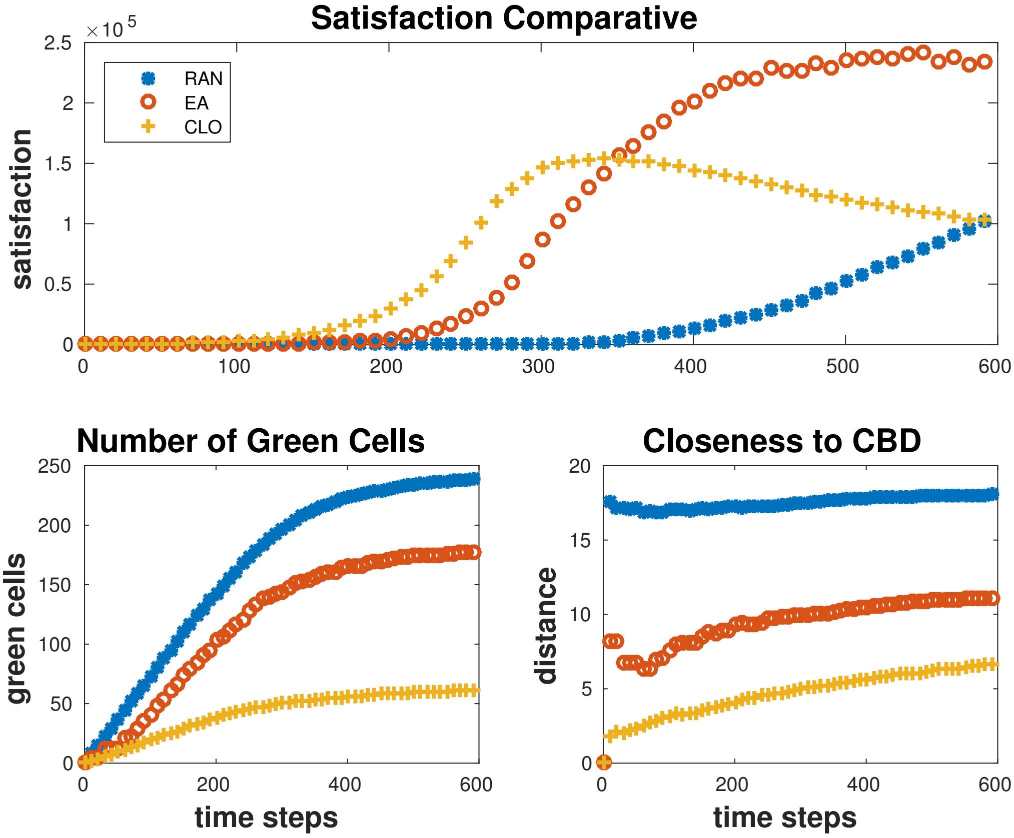



Figure 5.10: Results from scenario 3 (several CBDs and dynamic non-urban prices). The figure shows three different areas where the three approaches are compared: Satisfaction (top), number of cells protected (bottom left) and closeness to the CBD (bottom right). RAN approach is plotted in blue, EA in red and CLO in yellow.

In terms of the number of cells and the closeness measurement, EA manages to place its cells more conveniently, closer to the results achieved by CLO, at the expense of reducing the total number of cells in comparison with RAN. Finally the total amount of cells that each algorithm manages to buy is better for RAN and EA and steady in the case of CLO which is able to conveniently place its cells even if its total number is inferior to the scenario 2. The shape of the function that counts the number of green cells within the policy is similar to the previous scenario, which may mean that this factor is highly influenced by the nature of the rural prices (constant or dynamic).

| TIME STEPS | EA PERFORM. | RAN PERFORM. | CLO PERFORM |
|---|---|---|---|
| **50** | 0.00 | 0.00 | 202.57 |
| **100** | 99.80 | 0.00 | 1466.55 |
| **150** | 672.68 | 0.00 | 5424.48 |
| **200** | 2691.48 | 0.00 | 17139.55 |
| **250** | 11780.92 | 0.00 | 50052.82 |
| **300** | 47890.34 | 23.85 | 117803.02 |
| **350** | 121118.02 | 1430.46 | 151775.07 |
| **400** | 180117.40 | 8049.10 | 150859.87 |
| **450** | 214605.00 | 20455.57 | 139235.93 |
| **500** | 229921.00 | 39197.44 | 126713.13 |
| **550** | 235827.40 | 65507.47 | 114784.53 |
| **600** | 234665.60 | 93327.42 | 105446.80 |

Table 5.3: Numerical values of the EA, RAN and CLO algorithms' performance in terms of the satisfaction achieved by the urban population and measured by the fitness function during the complete time horizon of the simulation (data in line with Fig. 5.10 corresponding to Scenario 3)

## 5.6 Execution of the Offline Policy

A second mandatory step in the generation of an offline policy after its generation, consist of the execution of the plan created beforehand by the policy constructor component. In this phase, the level of validity of the generated policy is also checked.

A disadvantage of placing the unique policy execution after the planning phase is that some of the expectations and assumptions taken by the policy constructor may not be fully accurate. This may lead to certain selections of cells not being transformed as a result of a lack of budget compared with the real price of the parcel, or due to incompatibilities in the objective land class if this parcel is already urbanised. The test component executes the policy and checks the validity of a given EA solution. In order to do that, the present framework uses an independent simulation of the city where some data is gathered to provide information about the quality of the EA solution such as the amount of urban inconsistencies and the final satisfaction achieved by the solution.

When an inconsistency occurs in any of its types, the candidate cell to be protected at this time is rejected and no contribution to the final satisfaction is added. Its budget, in turn, is not spent and it accumulates for future purchases.

As a consequence, for every inconsistency found in the EA solution, the algorithm reduces its final satisfaction in comparison with the expected value calculated in the construction of the policy. Figure 5.11 shows the amount of urban inconsistencies found in relation to the number of simulations collected from the surrogate model.

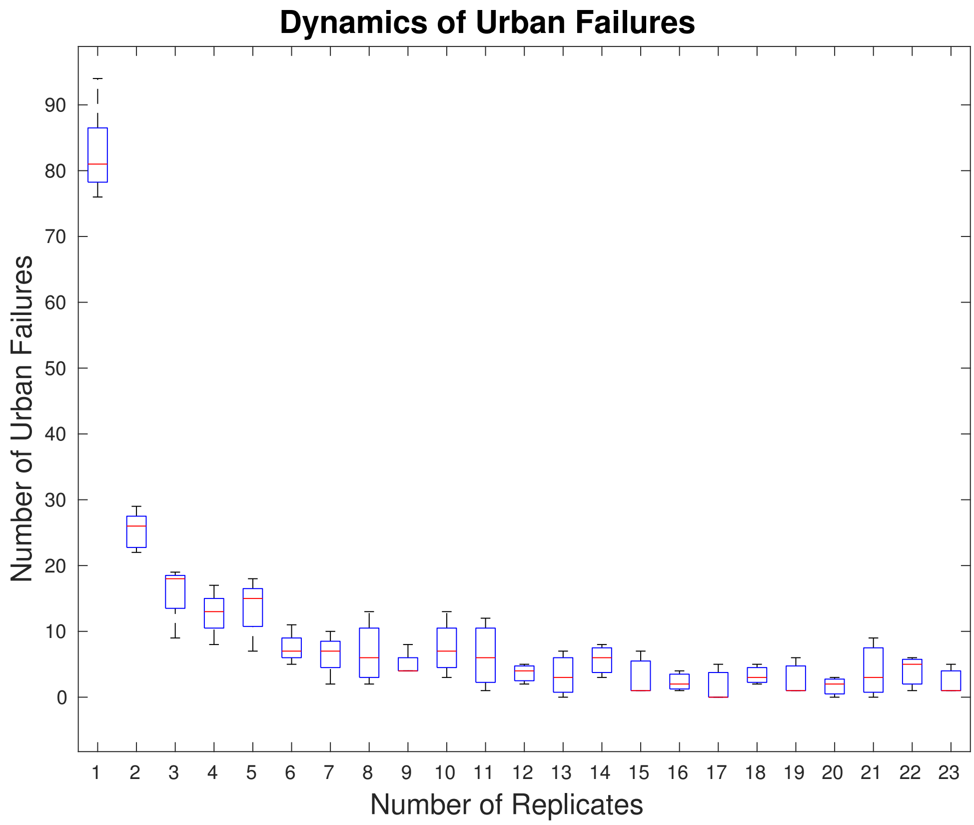



Figure 5.11: Evolution of the number of individual cells that are marked as an urban inconsistency as a function of the number of times the sampling process was carried out. Each box visually represents the number of urban failures from each specific number of samples. Averaged values are represented by the red lines and limits of the boxes depict maximum and minimum values gathered for each sampling size.

This plot is the result of gathering the number of inconsistencies that occur when the policy is repeatedly executed, concretely five times using the scenario two. The goal of this sample process is to gather enough data from the surrogate model to allow the EA to deal with the uncertainty derived from the unknown spatial evolution of the different land-use types during the creation of the policy. From the slope of the resulting function, it can be concluded that it is not necessary to gather a huge amount of data in order to achieve consistent results, due to the little variation in the number of failures ($\pm 5$). The slope of the graph shows similarities with an exponential decrement function in the number of rejections with respect to the number of simulations. By means of the analysis derived from the shape of this function, it can be concluded that it is not required to collect the data a large number of times to achieve accurate results. Concretely, from simulation number five onwards, the number of rejections are near steady and close enough to zero to be considered a robust method.

It is also noticeable that the model cannot eliminate the existence of inconsistencies in the execution of the policy, even if the number of replicates where the data is gathered increases. This is a consequence derived from the fact that the future cannot be completely predicted and the model always evolves differently in each execution. It is important to mention that as the urban model is used as a surrogate source of information, the number of failures is highly linked with the variability of the model, which means that it is a factor that has to be studied every time this EA strategy is applied in a different urban model. It is important to recall that these data come from non identical runs of the model, so this is a significant measurement of the behavioural consistency of the system.

The results shown in Fig. 5.12 correspond to a visualisation of the two elements that may produce local infeasibilities in the execution of the policy: the budget and the urban dynamics. A failure due to the budget is triggered when the cell assigned to be protected by the policy constructor (the EA optimiser) has not been urbanised yet, but the price is higher than the current available budget. Additionally, an urban failure occurs when the candidate cell has been previously urbanised, independently of its current price.

After the testing of the policy is finished, the module retrieves the total amount of failures provoked by both constraints. Afterwards, the process is repeated multiple times, where different policies are generated by each of the scenarios considered. Finally, these values are averaged over 10 runs for all the scenarios analysed.

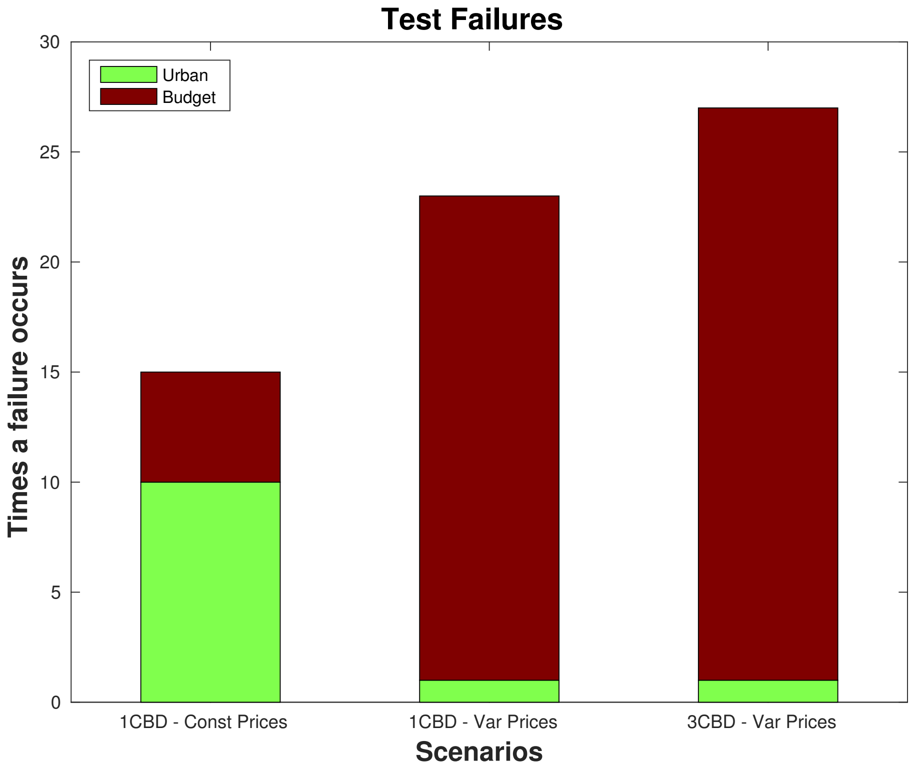



Figure 5.12: The graph shows a visual representation of the different types of inconsistencies that may be generated in the execution of the policy. Concretely a comparison can be made by measuring the areas generated by the number of failures caused by the budget (in dark brown) and by the amount of urban incompatibilities (in green) occurred during the execution of the different scenarios, each represented by an individual bar.

The study of the system can be performed under the assumption that the amount of failures of each type is in line with the configuration of each scenario, assuming a similar behaviour for equal level of complexity. This could imply that the amount of errors in the policy, caused by the lack of accuracy in assessing future urban dynamics, is similar in the first and the second scenarios, where the city was configured with a single CBD. However, this is not reflected in the gathered data. In fact, the general characteristic derived from the comparison of both quantities is that the offline EA approach found much more problematic to deal with the budget constraints than forecasting accurately the urban dynamics. Additionally, scenario 2 and 3 are almost identical in terms of urban failures, where the change to a higher number of CBDs seems not affecting the system. As a result, only in the case of the first scenario, where the urban prices are constant, the amount of failures is significantly lower and the contribution from the urban development is contrarily superior.

With respect to the number of failures due to the lack of financial resources, it can be seen that the system is more sensitive to this parameter and, consequently, the execution of the policy is likely to discard a higher number of cells due to this factor, especially in scenario numbers 2 and 3. Since the nature of the non-urban prices is different among scenarios, this seems to affect the number of inconsistencies in different ways. In the first scenario, these values are constant throughout the planning horizon. This deterministic behaviour would imply the generation of zero infeasibilities, since the system is always able to assess how much it should invest in each patch of land. However, the constant characterisation of the prices does not mean that they are homogeneous. Prices are based on two different types of land: forest and agricultural, each different in nature and subsequently in price, see table 3.3. The urbanisation causes an ecological degradation effect that would change forest cells into agriculture the closer the city is from them. This means that some cells could change the price during the simulation and a small number of infeasibilities can occur if the system tries to buy an agricultural cell that is, in fact, a forest land and hence, more expensive.

On the other hand, scenarios 2 and 3 lead to a large number of infeasibilities due to the fully dynamic nature of the rural prices. This factor may change, not only because of the type of land-use that they represent and the ecological degradation process linked with the urbanisation, as in the previous case, but also because the cells closer to the peri-urban areas of the city are significantly more expensive than the rest. These cells vary according to the continuous expansion of the city.

The amount of failures in assessing the price of the patches of land seems also being affected by more complex urban layouts, since this quantity in scenario 3 is higher than in scenario 2. This is caused by the cells located closer to the peri-urban areas of the CBDs. The amount of cells changing their state is higher when three CBDs are interacting than with a single central core. This difference influences the rural prices by speeding up the amount of cells in the grid that severely changes in comparison with the outer cells of the lattice, that remain more stable during most part of the simulation. Faster changes are more complex to assess and hence, the amount of failures increases also due to this factor.

## 5.7 Discussion

CLO is a very smart greedy strategy for simple urban planning problems where the allocation of facilities is conditioned by the fact that the population tends to concentrate close to the CBD. Following the experiments, it is concluded that the best results from this baseline were achieved when the model is characterised by constant rural price gradients, as in the case of Scenario 1, or when the allocation of green areas can be performed with unlimited resources, and hence there are no constraints in terms of budget. CLO can also accomplish a reasonable level of service when the growth speed of the city is slow and enough budget can be accumulated in order to buy the most expensive non-urban cells, which are the ones located closer to the boundaries of the city. This is concretely the reason behind the success of CLO strategy in a very short-term time horizon where the amount of available budget allows the complete provision of green areas within the city. However, in real world cases, when monetary resources are scarce in comparison with the price of the land, it is required to carefully plan and foresee future necessities of the population with the use of more intelligent approaches.

Because of this factor, the CLO heuristic shows the worst results at the end of the simulation. The total amount of cells protected are lower and concentrated mostly around the city centre, neglecting to give proper services to the outskirts of the city. At the moment funding starts to be insufficient, there are not enough resources to buy new cells in the peri-urban areas of the city, and then the strategy reduces its performance. Additionally, when more flows of population move from the CBD to outer areas of the city, the satisfaction achieved by CLO is reduced drastically getting even closer to the performance of a simple random approach.

Table 5.4: Characteristics of the execution of the Evolutionary Algorithm

| Scenario | Generations | Mutations | Worst | Best | Mean | SD |
|---|---|---|---|---|---|---|
| 1CBD - Const Prices | 2045 | 1169 | 2583 | 2613 | 2601 | 13.764 |
| 1CBD - Var Prices | 1982 | 1565 | 1262 | 1271 | 1269 | 6.578 |
| 3CBD - Var Prices | 1507 | 1144 | 1192 | 1197 | 1195 | 3.838 |

General characteristics of the evolutionary algorithm among scenarios. First column shows the number of the scenario where these data was collected. The second column describes the number of generations that the algorithm needs to run until achieving convergence. Column three depicts the total number of mutations where the algorithm was able to improve a member of its population. Columns four, five and six illustrate the difference between the members of the population of solutions by representing the final worst, best and average values respectively of their fitness function. Finally the standard deviation from these values is calculated.

For complex scenarios like 2 and 3, the EA algorithm outperforms both baselines during the entire simulation. This approach also achieves a steady level of service even when the resources from the budget are insufficient. This advantage is an effect of the design of the fitness function that takes into account the satisfaction of each green area from the moment that the patch of land is purchased to the end of the planning period.

In summary, the CLO baseline, that always allocates the urban parks as close as possible to the most populated areas, is very powerful for simple scenario configurations when the distribution of non-urban prices is not characterised by drastic gradients. EA, on the contrary, is able to provide more efficient solutions in the long-term for the entire population. It can also deal better with variations in the distribution of the population and it can tackle more satisfactorily scenarios configured with heterogeneous peri-urban prices and more complex urban structures.

These results are also in line with the conclusions held by Golovin et al. (2011) for problems that can be characterised by an adaptive submodular property like this given case. An intelligent greedy approach like CLO is able not only to compete with an EA approach producing better policies in terms of performance, but also generates better results under certain circumstances like the ones described previously in this section. Apart from that, it is also worth noting that CLO, as a greedy approach, is able to generate its results much faster than the EA algorithm due to the simplicity of its strategy.

Finally RAN solutions, due to the stochastic selecting mechanism, spread their protected cells more homogeneously and scattered throughout the surface under consideration. This strategy does not take advantage of the reduced prices around the city at the beginning of the simulation. In turn, when prices are significantly higher and it is more difficult to purchase any new patch of land, the RAN approach manages to give a reasonable service to the outskirts of the city, in contrast to the other two approaches that are more sensitive to the significant increment in prices occurring at the end of the simulation.

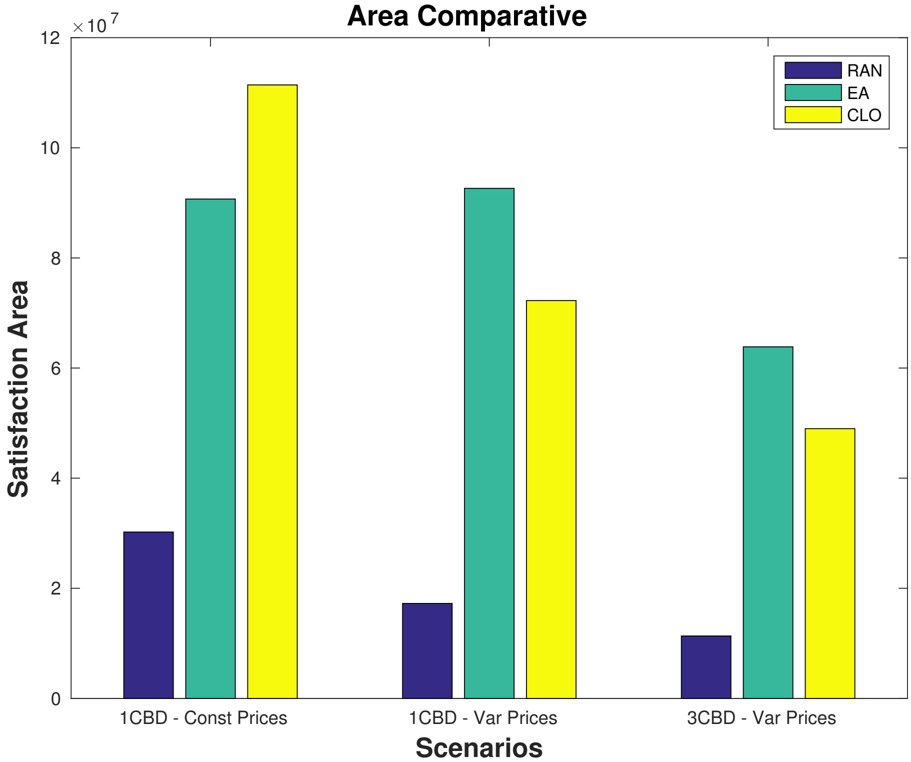



Figure 5.13: This bar plot illustrates the evolution of the satisfaction of the different optimisation approaches through the three designed scenarios. Each group of bars measures the area generated by the fitness function that assesses the satisfaction achieved by the population of the city. In dark blue is represented the area covered by the random approach RAN, in light blue the one generated by the EA algorithm and in yellow the same results for the CLO baseline.

Fig. 5.13 shows the results achieved by the fitness function, focusing on the area of the surface generated by the curve that describes the satisfaction in each time step instead of its shape. This area will be calculated using trapezoidal numerical integration techniques (Yeh and Kwan, 1978). From this perspective we can have a global assessment of each strategy, allowing us to measure the behaviour among scenarios and to be able to compare the three approaches from a more general point of view. This evolution in satisfaction could give us information regarding the effect that the changes in complexity have on the algorithms.

From a preliminary visual inspection, if each maximum point of the bars generated by the algorithms in each scenario are connected as a curve, it can be observed that CLO bars, in yellow, represent a function that is similar to a decreasing linear function. RAN baseline in dark blue depicts a function that can be generalised to be closer to a power-law distribution and finally EA forms a function that can be closer to a negative binomial distribution distribution. These statements are based on the idea that the order of the scenarios is a measure of the complexity of the system. The behaviour of these functions has to be taken as a preliminary conclusions and further analysis should be carried out in order to confirm their validity. The analysis should include more configurations to add more points to each function and the inclusion of other factors that can potentially add significant complexity to the system.

It is also worth noting that there is a swap in performance between the CLO baseline and the EA algorithm in scenarios 1 and 2, where EA outperforms CLO. However, this tendency is not maintained between scenario 2 and 3, in which CLO get closer in terms of the area covered to the EA approach. This may imply that the change in the nature of the urban prices benefits the EA, creating a search space where EA can exploit more its potential meanwhile the spatial complexity of the layout of the city does not benefit the algorithm excessively.

From the observation of the behaviour of the fitness generated in each of the different scenarios in Fig. 5.13, it can be observed that the values gathered from the CLO baseline compared to the developed EA drops when higher complex scenarios have been taken into consideration, especially when it is added another factor of complexity to the model, namely the non-deterministic rural prices (scenario 3). However, the different spatial configurations resulting from including multiple

CBDs into the model, does not give rise to a clear benefit over the EA algorithm. Subsequently these limited results show that the non-adaptive EA approach is a suitable optimisation technique for location allocation problems with complex dynamics and constraints, but less suitable in scenarios where the submodularity property applies, in which case a greedy algorithm may be preferred.

## 5.8 Conclusions

In this chapter, it was tackled the task, faced by many urban planning departments, of green space allocation over time in growing and densely populated urban environments. This was done by using the urban growth simulation framework and the optimisation strategies that were respectively described in previous chapters.

The problem investigated was formally expressed and an important property of the system into consideration was explained. Adaptive submodularity has a high impact on the final performance of the algorithms. This property implies that, under some circumstances, an intelligent greedy algorithm can overperform a more complex approach. To study how the developed EA behaves a series of scenarios were defined, varying the complexity of the topological configuration of the lattice, and the nature of the non-urban prices.

Different optimisation methods were then tested against the EA: two rule-based baselines, one which chose cells closest to the CBD (CLO) and a random baseline (RAN). The CLO baseline was found to achieve very good results in terms of estimated satisfaction of the residents, but at the expense of significant budget spend. Over simulated time, this effect is exacerbated, and CLO fails to compete with EA, especially in the long term, where the population settling in the outskirts of the city cannot be adequately served by green space provision. Meanwhile, experiments showed that the results achieved by our EA strategy begin to outperform the greedy strategy with increasing model complexity, for example involving a realistic city configuration with multiple cores, multiple housing price gradients and occasionally turbulent market dynamics.

It was argued that this higher complexity model more closely resembles real world conditions, and therefore we can conclude that EA approaches are better suited to inform real-world decision making. These conclusions are limited by the proof of concept nature of this framework and the very small number of scenarios analysed. Further systematic experiments would be required to achieve more consistent results.

In the next chapter it is compared this adaptive EA with an 'online' version. Both approaches will investigate the same problem of generating open space plans by solving a SDMP in a complex, stochastic urban scenario. In this case, it is analysed their performance and computational time, and it is also presented a qualitative comparison between the types of policies generated by the two methods and the topological structures generated in each case.

# Chapter 6

# Online Evolutionary Algorithms for Planning

## 6.1 Introduction

Urban-planning authorities continually face the problem of optimising the allocation of green space over time in developing urban environments. The problem is essentially a sequential decision making task involving several interconnected and non-linear uncertainties, and requires time-intensive computation to evaluate the potential consequences of individual decisions. In this chapter, the application of two very distinct frameworks incorporating evolutionary algorithm approaches for this problem are explored: (i) an 'offline' approach, in which a candidate solution encodes a complete set of decisions, which is then evaluated by full simulation and (ii) an 'online' approach which involves a sequential series of optimizations, each making only a single decision, and starting its simulations from the endpoint of the previous run. The outcomes, in each case, in the context of a simulated urban development model and compare their performance in terms of speed and quality are studied. The results show that the online version is considerably faster than the offline counterpart, without significant loss in performance.

## 6.2 Definition of the problem

The aim of green space planning is to select an efficient sequence of green space facilities over a defined period of planning time. This task can be viewed as a SDMP where a sub-optimal finite set of $N$ sequential alternatives has to be found in a discrete-time system. In a green space planning scenario, each of these $N$ alternatives can be seen as an embedded LA problem. LA problems, first described by Cooper (1963), can be defined as the optimal placement of a set of new facilities or units in such a way that a series of goals is met.

EAs have been successfully used to solve complex spatial problems (Pukkala and Kurttila, 2005; Lu and Eriksson, 2000) in discrete decision spaces (Davis, 1991; Goldberg, 1989). However, their performance under uncertainty has been questioned (Rieser et al., 2011; Wu et al., 2006) since a simple EA has insufficient information to deal directly with uncertainty.

To equip an EA to solve this kind of problem, it is necessary to engineer appropriate mechanisms which help maintain suitable search progress without being misled by noisy decisions, or unacceptably decelerated by time-consuming simulations. Existing methods that can be applied to evolutionary systems to cope with such circumstances include tools such as noisy fitness, fitness approximation and dynamic fitness functions. In (Qin et al., 2010) a GA-Aided stochastic optimisation model is applied to cope with the uncertainty related to the study of air quality in urban areas. In contrast to probabilistic approaches, (Wang and Yang, 2009) resort to local search techniques to overcome the uncertainty generated by the ageing factor presented in many engineering problems. Following a similar approach (Wang et al., 2013) successfully applies a variant constrained multi-objective EA in a simulated topology and shape optimisation problem under uncertainty.

To further test the potential of evolutionary techniques in this area, the stochastic model is configured to use a topologically non-trivial city with several CBDs and different price gradients (see Fig. 3.1). This type of arrangement complicates the search and decision space and provides a scenario where EA can better shows its potential of coping with these additional complexities (Pukkala and Kurttila, 2005; Vallejo et al., 2015).

Price gradients can be divided into two different dynamics according to the type of land they represent, broadly subdivided into rural and urban. Urban prices are higher the closer the patch of land is to a CBD, and also depend on household agents' preferences and on the current level of demand for this particular parcel of land. On the other hand, rural prices are different if the corresponding area is classified as a forest or as an agricultural cell. They are also influenced by the distance to the closest peri-urban area, due to the impact on the expected profit derived from its urban transformation. Rural land, located in the surroundings of the borders of the city, achieves a peak in prices, decreasing from this point with distance (Plantinga et al., 2002). Meanwhile, the continuous urban growth expansion narrows the amount of available land, and this diminishing supply increases the price of the remaining land as time passes.

The provision of green services is performed following a covering model (Toregas et al., 1971). The principle of this approach is to maximise the number of users who are located relatively close to the defined type of service, in this case green areas.

### 6.2.1 Experiments Conducted

This chapter investigates the viability of two distinct approaches using evolutionary algorithms as alternatives for generating open space plans by solving this SDMP in a stochastic urban scenario. In what we call the 'offline' approach, a complete plan is proposed at the outset; that is, at time=0, it has already been decided which parcels of land are slated to be purchased at all future time steps (if it turns out to be feasible in the ensuing simulated circumstances). In contrast, the 'online' approach makes its land-acquisition decisions one at a time, each time benefiting from the reduced uncertainty arising from the previous step. Both approaches are empirically evaluated and compared in response to a set of physical and ecological constraints using the same software structure for all methods.

To measure the suitability of each approach, three criteria have been simultaneously taken into account: the computational time used to run both approaches, the objective function value achieved during the length of the planning period and the spatial patterns generated by the final layout of the protected cells selected by each approach. The implementation of the offline policy is identical than the one used in previous chapter. Same techniques to deal with the uncertainty are used in both approaches. For each of the algorithms, different purchasing strategies are considered. A purchasing strategy can be seen as a decision making process related to when it is better to save the received financial resources and when it is worthy to invest them in a new patch of land. In this regard, the offline approach will include the ‘fix’ and the ‘stochastic’ strategy and the online algorithm will consider a ‘threshold-based’ strategy.

## 6.3 Online Optimisation Procedure

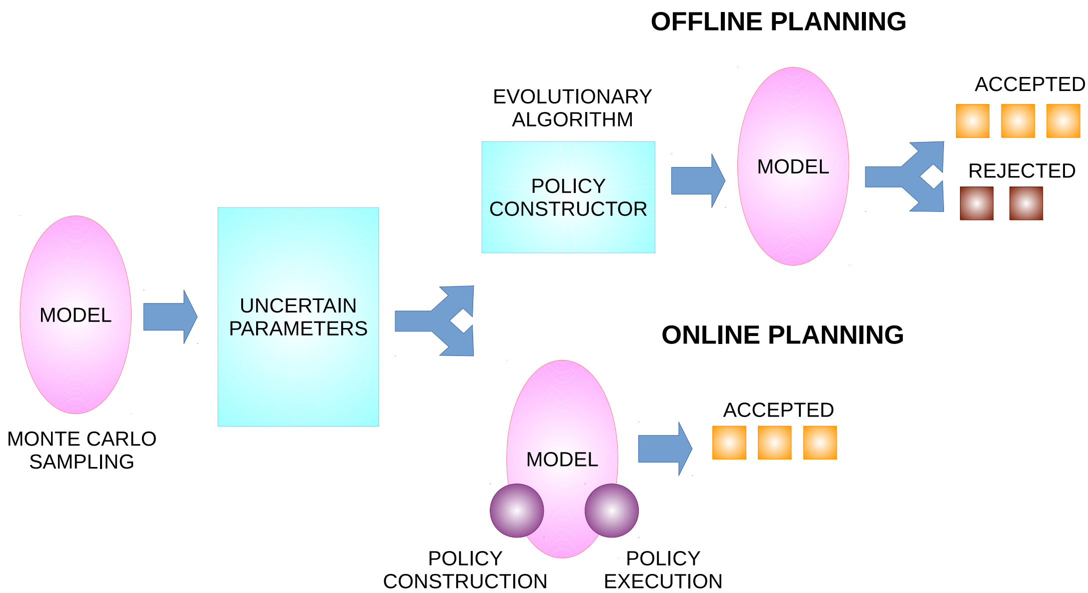



Figure 6.1: The schema of all the elements included in both planning processes is depicted in this figure. The previous generation of uncertain parameters, how these different components are linked with the hypothetical urban model and the different nature of the results in each approach are also included.

Based on this evolutionary strategy, the proposed optimisation framework includes a variety of components as can be observed in Fig. 6.1 and Fig. 6.2. In these figures, two different EA workflows for each scenario, an 'online' and an 'offline' approach, are described. Both approaches share a common budget and an initial environmental scenario. By means of Monte Carlo sampling techniques, the offline approach will gather the required information about the most probable population density and its distribution, the urbanised areas and the rural prices for each time step considered in the simulation. The online counterpart only needs the population information to calculate the fitness function. Based on that data, the optimisation procedure depicts the adaptive and non-adaptive nature of the corresponding online and offline approaches. Finally, the generation of the final decision alternatives is performed which describe the concrete policies resulted from the process.

The main goal of the optimisation process is to select a series of patches of rural land, within a determined time horizon, to be transformed into green areas as the city develops over a number of years. The general objective pursued with this purchasing policy is to attempt to ensure that each land allocation decision fulfils both the present needs of the urban population, along with estimated needs of the larger population as it develops in future time steps.

The optimisation process starts in both cases with the definition of two initial constraints that are imposed on the algorithms at the beginning of the simulation: a common *budget* which limits the acquisition process, and an environmental layout whose values are attached to every cell of the grid. In Fig. 6.3, an example of this ecological lattice is shown. From a general point of view, this lattice shows the ecological value degradation effect caused by urban development in a city of three CBDs. The small green areas within the urban cores depict the protective impact of allocating green parks in the city.

In both cases the constraints are generated in advance by a uniform random process. The budget takes values in fixed intervals for the entire duration of the simulation. The ecological values measure the natural resources richness of the land and determine the rural land type: cells with high 'eco' values are classified as forest, while rural cells with low eco-values are classified as agricultural. The way these values are generated at the beginning of the simulation is described in Algorithm 3.3.

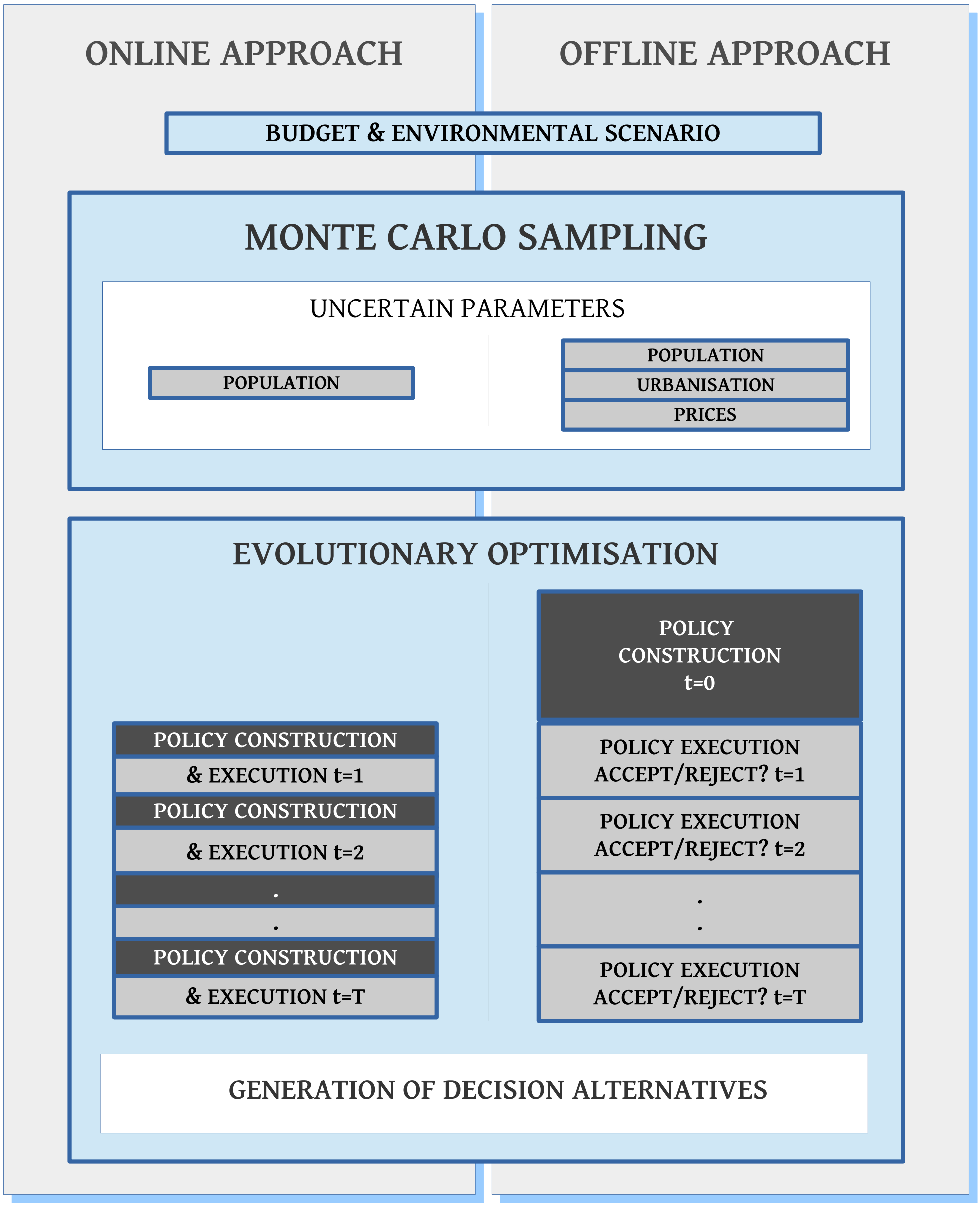



Figure 6.2: The online & offline optimisation workflows are depicted highlighting the characteristics that these two strategies have in common.

The ecological value of a cell, and hence its land-use classification, is also influenced by the values assigned to the cells in its neighbourhood. This leads to a dynamic diffusion/feedback process within the model, which is capable of mimicking the ecological degradation process provoked by urban expansion over the remaining open space (Alberti and Marzluff, 2004). Both of the initial constraints (budget and environmental layout) directly affect the land purchase mechanism. The budget limits the locations and amounts of cells that the system is able to acquire, and the ecological values influence rural prices and the non-urban land-use type.

In terms of their formulation, the two algorithms vary in line with the nature of

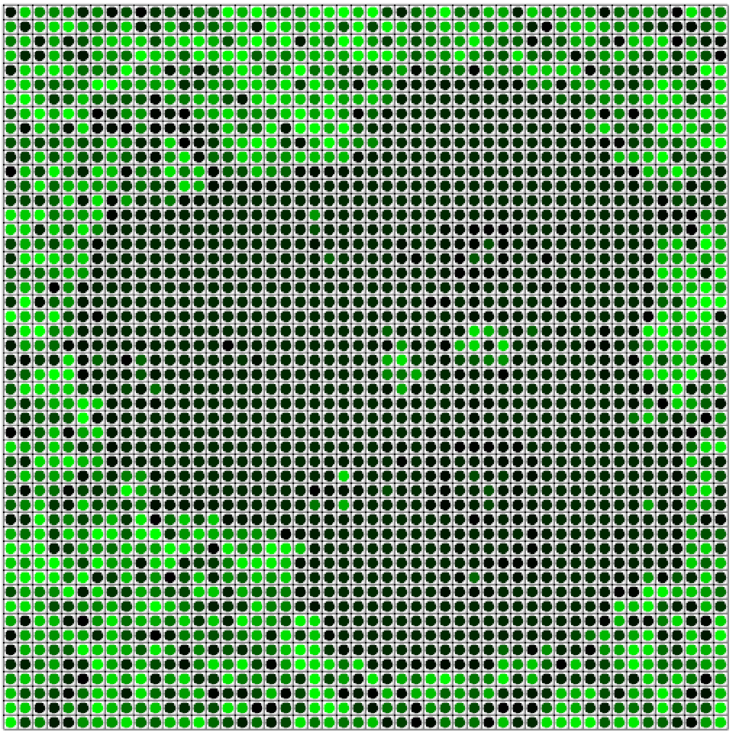


Figure 6.3: Visualisation of a determined configuration of ecological values linked to each cell where black cells represent very degraded areas and green cells describe rich environmental zones. The range of green tones represents intermediate states, with brighter colours depicting more valued areas.

their overall approaches, which, respectively, are adaptive and non-adaptive. The main difference between them is that during the execution of the online approach, local policies are computed at each decision step; whereas, in the offline algorithm, all policies are decided and fixed at the beginning of the simulation.

### 6.3.1 The Offline Algorithm

The implementation of this algorithm was explained in detail in previous chapter. Summarising briefly, the algorithm starts with the random generation of an initial population of potential solutions using different seed values. A relatively small population size of 25 is used for both approaches, following preliminary work in which we concluded that this enabled both algorithms to converge relatively quickly, and considering each algorithm individually, was not surpassed in solution quality by larger population sizes. Each chromosome stores the information needed for encoding a complete solution to the problem. The construction process has to satisfy some initial constraints in terms of budget and land availability. Protection is not allowed in urban areas and budget must be strictly positive at any time. Further characteristics of this approach can be seen in previous chapter.

### 6.3.2 Purchasing Decisions in Offline Planning

Due to the time-dependent nature of the budget, the monetary resources available at a given time $t$, are partially the result of the previous land purchase history. Hence, the selection of the time steps, where the land acquisition should be carried out is one factor that severely influences the effectiveness of a given policy. There are different criteria that could be applied to decide if it is a good choice to spend the current budget in a given time step or whether it is more convenient to accumulate it for the future in order to select a more appropriate and expensive patch of land. In the offline algorithm two different approaches were implemented to solve this problem:

- **FIXED**

  Every time the budget is bigger than the cheapest cell in the grid the system tries to make a purchase. If in 100 attempts the algorithm is able to find a suitable selection of cells, the acquisition is performed. In general this approach can only afford to buy a reasonable amount of cheap cells located far from the CBDs, due to the fact that the continuous acquisition of land refrains the strategy from being able to save enough financial resources to buy other kind of patch of land.

- **STOCHASTIC**

  In this case, purchase decisions are limited not only by the available funds, but also by a variable which restricts the purchases to the 70% of the times that these actions can be potentially successful. This value was selected by a test and error process, searching for the final best values of the fitness function. The strategy allows the system to focus on a more qualitative type of cells than the previous approach at the expense of dropping the total number of cells transformed into green spaces.

Both methods treat the entire set of decisions as equally important. However, earlier purchases are more critical than the later ones, due to the fact that at the beginning of the simulation, when the city is small in size there are promising empty areas close to the CBDs that will soon be urbanised. These areas will potentially be highly populated in the near future and the amount of satisfaction collected will be much higher than in zones closer to the outer limits of the grid.

### 6.3.3 The Online Algorithm

The online approach performs a series of optimisation runs, one for each time step, to support the decision at that time step. In contrast to the offline approach, each of these individual optimisations requires a much simplified chromosome $Cr$, which only needs to represent the current state of the system since it can exploit the complete knowledge of the current belief at any moment. Consequently, its structure is reduced to the following information:

$$Cr_t = [c_{(x_1,y_1)}, c_{(x_2,y_2)}, ..., c_{(x_m,y_m)}] \tag{6.1}$$

where each cell $c_i$, spatially located in a different pair of coordinates $(x, y)$, belongs to the set of $[0, m]$ cells selected for being transformed into parks in the present time step $t$. Notice the variable length of the chromosome. Another difference from the offline approach is that, it is not necessary to accumulate any remaining budget because the available funding, shared by the entire population, is known at each moment. This property of the budget allows the system to completely remove this source of uncertainty.

Regarding the mutation operator, the online approach uses the same mechanism explained in Sec. 5.4.2.3, but with a different level of granularity. Instead of searching for an individual selection $\beta$ to mutate amongst the entire set of selections, as it was done in the offline algorithm, the online mutation generates a complete new individual $Cr'_t$, where $Cr_t \cap Cr'_t = \varnothing$.

Contrary to the previous approach, all the selected cells included in the population at any time step are feasible. The optimisation module is aware of the current state of the system and the entire set of chromosomes constructed and evolved under these circumstances are valid and can be checked at the time the policy is decided.

Due to time restrictions typical of practical online planning, the online stopping criterion is halted prematurely, reducing the maximum number of generations that the algorithm is allowed to run without improving to 50 and the total number of generations to 450.

The pseudocode of the selection of green cells by the 'Online' algorithm is the following:

**Algorithm 6.1** Online Selection of Green Areas

```
global variables
    TOTAL_TICKS
end global variables
require currentTick
procedure updateAggregate(int currentTick)    ▷ The system evaluates possible acquisitions in each time step
    if currentTick == TOTAL_TICKS then
        RunEnvironment.getInstance().endRun();
    else
        searchGreenSpace(currentTick);
        writeSatisfaction(currentTick);
    end if
end procedure
require currentTick, solution[], size
require coodX, coodY, cell
procedure searchGreenSpace(int currentTick)
    solution = municipality.runGA(currentTick);
    size = solution.length/2;
    if solution.length != 0 then
        for i=0;i<size;i=i+2 do
            coodX = Integer.parseInt(solution[i]);
            coodY = Integer.parseInt(solution[i+1]);
            cell = GreenArea.getCell(coodX, coodY);
            if cell==null then
                System.err.println("Cell not found when trying to protect");
            else
                cell.setState(CellState.PROTECTED);
                protectedCells.add(cell);
                nonUrbanCells.remove(cell);
            end if
        end for
    else
        System.err.println("Solution length equal to zero");
    end if
end procedure
```

In the procedure *updateAggregate* significant variables are *currentTick* and *TOTAL_TICKS* which represent the current time step of the system and the total length of the simulation respectively. Additionally, for the *searchGreenSpace* procedure, the algorithm uses the following variables: *solution* that it is the list of cells selected in this time step by the *municipality*, *size* that represents the number of cells to be protected, *cell* which stores each of the cells to be acquired and their coordinates in the grid, depicted by *coodX* and *coodY*. The time complexity of this algorithm depends on the complexity of the function *runGA*.

### 6.3.4 Purchasing Decisions in Online Planning

As a result of the simplified nature of each online individual, the question raised is if it is possible to implement a more efficient mechanism to discern whether it is convenient to look for a feasible candidate patch of land to purchase, which means to evolve the current EA population in this turn or save the resources for the future.

One possible way to answer this question is by analysing the nature of the search space and the way it is constrained by external factors at the current point in time. The method used is referred here as a *threshold-based strategy*, and it involves recording the mean fitness of the initial population of random solutions at the current time step. The quality achieved at this initial point provides a useful hint about the potential capacity for improvement by evolution, subject to the current budget and land price constraints. Such use of information retrieved in the first generation was also explored in Ratle (1998) to create a global approximation of the fitness function.

Purchasing decisions are therefore taken by comparing the averaged values of the fitness function of the initial population with a pre-defined fitness threshold. If the fitness threshold level is not met, it is assumed that, even by the use of an evolutionary process, the algorithm will not be capable of finding enough attractive choices in this current time step, presumably due to the gap between the current budget compared to the current prices.

In the case that the initial population check passes the threshold, and consequently, we evolve a population at the new time step, a second fitness threshold comes into play. This time the new threshold is based on the averaged values of the fitness function of the evolved population compared to the average value of the fitness function of the initial population. If the second fitness threshold is not reached, we take the view that purchases made in this time step are likely to be suboptimal due to a lack of evolution, and may cause a cascade of further difficulty in future time steps through poor use of budget in this round. Not reaching this second fitness threshold, therefore leads to rejection of this result, and we reinstate the budget and proceed to the next time step, where the current budget will be accumulated. Both thresholds were defined and calibrated by empirical observation of the algorithm's behaviour in preliminary work.

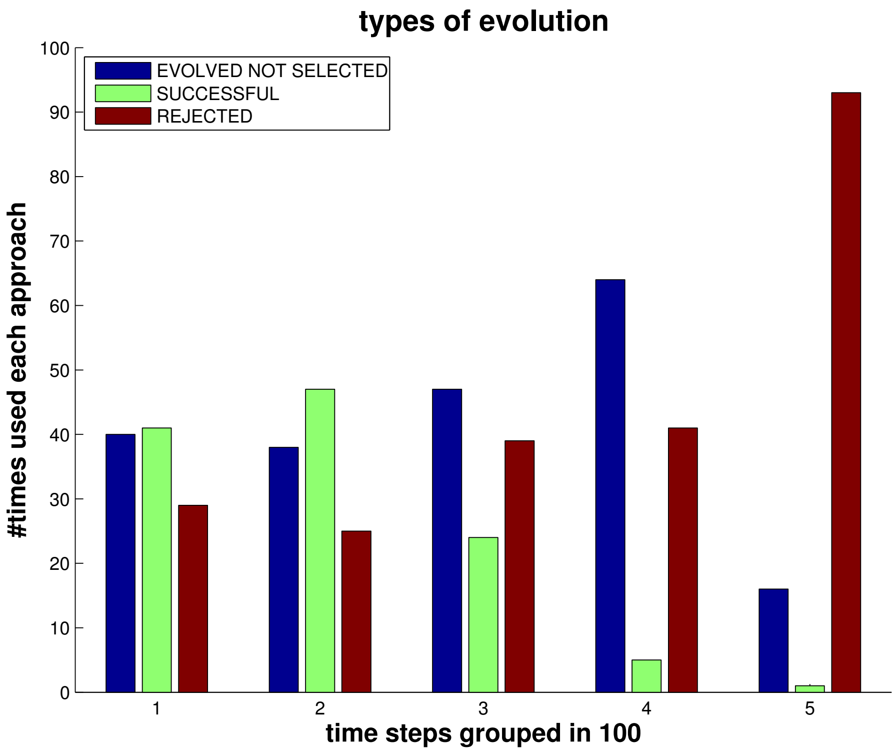



Figure 6.4: Amount of different type of evolutions of the population according to the nature of the search space in each concrete time step. Green values represent successful evolutions, where the policy was decided and executed and consequently a patch of land was protected. Red values represent generations with an insufficient level of quality in the initial population, where the population was not evolved. Finally, blue values illustrate promising generations in which the level of improvement was not enough to be considered as a suitable solution for being included into the final policy.

In Fig. 6.4, we depict the behaviour of the evolution resulting from the application of this strategy. Successful attempts peak at around 200 generations, where almost 50% of the time the initial population of solutions was accepted. From this point, there is a clear decline in successful attempts; at generation 400 only small gains are made in terms of patches of land selected and protected. The evolved and not selected set of solutions is a strategy executed around $40\% - 50\%$ of the times during the first half of the simulation, increasing the frequency up to more than 60% in the second third part of the simulation, decreasing drastically to less than 20% at the end. Meanwhile, rejected generations reach $20\% - 30\%$ in the first third of the simulation, climbing up to 40% in the second third, until, towards the end of the run, all generations are rejected, since rural prices are extremely high due to the lack of supply.

In terms of time consumption, the first threshold speeds up the algorithm, avoiding the exploration of certain generations meanwhile the second threshold wastes the time taken in evolving a population that perhaps is not going to be used within the final policy.

## 6.4 Evolutionary Algorithms and tools to deal with Uncertainty

In general, online approaches alleviate the computational complexity and the level of uncertainty by only considering the state of the present situation and the current time horizon, whereas an offline search would compute a large contingency plan prior to its execution, considering all possible situations that could occur (Ross et al., 2008).

A step-by-step planning strategy is significantly more flexible and it allows the online version to handle possible changes in the environment without extra computation. This characteristic allows planners to take the potential advantage of analysing and responding to the stochastic dynamics of the system. On the other hand, the offline approach can get benefit from, effectively, being able to look further ahead, acquiring at lower prices patches of land that are located in more remote areas. These locations, that currently are not very attractive for the online strategy, can be unaffordable when the level of interest increases.

An additional advantage of the modular structure of the offline EA is that it can be applied straightforwardly to any urban growth model, even if an increment in the complexity of the model may require substantially more computational effort than the one reported in this thesis. The model, in this case, would only need to read the output file generated by the offline policy constructor and applied it afterwards.

The online counterpart, on the other hand, would need to be integrated into the code of the objective urban growth model in order to receive the input information that characterise each individual realisation. Its applicability would generally require further coding in order to include the policy constructor considered here with the execution module, which is particular to each individual urban model. Hence, due to this continue feedback of information required from the urban model, the design of the online component cannot be independent.

### 6.4.1 Offline Constraints & Testing

When a sequential set of decisions has to be taken in advance with no information about the state of the environment at these times, highly constrained conditions can easily arise at the time the policy is implemented. In this regard, the offline algorithm ideally needs to receive as inputs accurate information about which cells are urbanised during the entire period considered and the dynamics of the rural prices. By using Monte Carlo sampling techniques analogously, as was previously done for the urban population, this information is collected. Finally this urban information will be used to constrain the set of available parcels of land. Meanwhile non-urban prices will limit the cells that the budget can afford during the simulation.

In Fig. 6.2 it is shown that the offline optimisation approach is divided into two main independent parts: the policy constructor implemented by the EA, and the policy execution which is covered within the code of the urban model (the CA). A disadvantage of placing the unique policy execution after the planning phase is that some of the expectations and assumptions of the policy constructor could be found to be wrong when facing real conditions. This may lead to certain selections of cells included in the planning that cannot be transformed into green areas due to a lack of budget compared with the real price of the parcel at this time, or also because of incompatibilities in the objective land class in the case the parcel is already urbanised. The policy execution, responsible for checking the final set of cells, is allowed to dismiss the candidate green areas which are incompatible with the real instance of the problem. Once the filtering is performed and the final set of cells is defined, the policy executor will calculate the real satisfaction of the offline algorithm based on this successfully applicable subset of cells.

## 6.5 Results & Discussion

The results shown in this chapter are the averages over 20 independent optimisation runs for each approach. The algorithms were executed in a Linux operating system with an Intel core i5-3210M processor and 8GB DDR3 of RAM memory and they were coded in Java using an Eclipse compiler and the open source software RS version 2.0 (North et al., 2005). RS is an agent-based modelling and simulation toolkit commonly used in the CA-ABM community. Previous publications have validated the software in different contexts (Parry et al., 2006; Griffin and Stanish, 2007).

The previously described hypothetical CA-ABM urban growth framework was used, with the main objective of making use of the complexity of its dynamics (Batty et al., 1999) in order to test and measure the suitability of EA techniques in a stochastic scenario under uncertainty in two different versions, the described online and offline approaches. The CA consisted of a two-dimensional regular lattice of $50 \times 50$ cells with i- and j-axes and stochastic transition rules. The model is configured to develop an urban area with three main CBDs growing in parallel with different price gradients.

### 6.5.1 Computational Time

The CA-ABM model, as previously mentioned, was implemented using the RS framework. This is an important aspect to take into consideration when the computation time of both algorithms is analysed. The RS framework requires a significant amount of computational resources to simulate the dynamics of the city, the representation of all its inhabitants and their interactions.

| Approach | PC (sec) | PE (sec) | TT (sec) | TR | #Gen |
|---|---|---|---|---|---|
| Offline | 41011.8 | 911.32 | 41923.12 | 2.145 | 1682 |
| Online | 19118.3 | 1604.8 | 20723.1 | 0.466 | 450 |

Table 6.1: Offline vs. Online results in terms of computation time and the number of generations.

The resulting time statistics are shown in Table 6.1. Column 1 *Approach* refers to the name of the approach considered. Column 2 (PC) *Policy Construction* represents the running time that both algorithms require to construct their policies without taking into account the policy execution. Column 3 (PE) *Policy Execution* comprises the implementation of the planning and the evolution of the rest of the model dynamics. Column 4 (TT) *Total Time* illustrates the total used processor time: which includes the previous policy construction time plus the policy execution. Column 5 (TR) *Time Ratio* shows the ratio of offline time to online, that is the quantitative relation between them, showing the number of times one value contains the other. Finally, Column 6 *#Gen* depicts the average number of generations used in each approach.

It is intuitively clear that offline approaches would need considerably more time to come up with a final plan in comparison with an online version of the same algorithm (Ross et al., 2008). In the present work, computer simulations show that the proposed algorithm exhibits the expected behaviour in terms of time.

It should be noticed that there are two main factors that can significantly affect the behaviour of the time consumed. Firstly the offline version can run alone in the system console using all the available resources, meanwhile the online version has to share memory and processor resources with all the infrastructure created by Repast Simphony and Eclipse. This means that at the same time that the EA population is evolving, thousands of objects which represent agents, cells and other type of elements of the urban growth model will typically coexist. This advantage is, however not enough to speed up the algorithm significantly. Secondly the number of generations that the online version is limited to evolve in comparison with the offline version aids the algorithm to show better time performance at the expense of decreasing the fitness of its final individual solution.

One drawback associated with online planning is that, in a practical setting, it would generally be required to meet real-time constraints; this means that the algorithm would need to greatly reduce the available planning time to satisfy the requirements of a real-time environment. For instance, this could be the case of robot motion planning or anytime learning scenarios (Gaschler et al., 2013; Vargas et al., 2014). However the system, in any of its versions, cannot properly cope with the requirements of a real-time system, because the nature of EA requires long computational times until convergence and it is not considered in general a suitable approach for these kind of scenarios (Ciesielski and Scerri, 1998) unless the algorithm is strictly time-bounded.

Other external characteristics that can be associated with the time calculation is the convergence rate and the internal design of the EA. Changes through better areas of the search space require time in terms of number of generations. As was commented previously, Alonso's model favours areas of the grid closer to the CBD, which are more and earlier populated. In the present offline implementation, since the crossover operator is dismissed, every generation the algorithm is able to modify a single allele of one individual of the entire population which means that the mutation rate $p_{of}$ can be at maximum:

$$p_{of} = \frac{1}{T}\frac{1}{N_{pop}} \tag{6.2}$$

where $T$ is the number of time steps of the simulation and $N_{pop}$ is the number of individual solutions evolving together in the evolutionary population. In turn, in the online approach the mutation rate, denoted by $p_{on}$, will change the information of one single individual in each generation:

$$p_{on} = \frac{1}{N_{pop}} \tag{6.3}$$

Following from that, the mutation operator in the online approach allows further jumps in the search space, converging faster than its offline counterpart. This aspect also means that it is more likely that the online algorithm falls into local minima. However, since the evolving time assigned is rather limited, a faster improvement of the solutions compensates for this effect, since in general chromosomes do not have enough time to properly converge.

In conclusion, the offline algorithm navigates the search space slowly, making a large number of small steps; it is essentially a highly exploratory and 'careful' approach, capable of finding near-optimal results, but generally requiring many generations to do so. Meanwhile, the approach used by the online algorithm is more an exploitative approach, taking a small number of large steps; this is a strategy that can provide excellent results in a short time, but generally runs the risk of not achieving the best results available.

### 6.5.2 Performance

In Fig. 6.5, the performance of both algorithms is visualised. Performance is calculated by measuring the satisfaction achieved by the population using the fitness function, see formula 5.11 on page 168. At this point it is important to recall that a person living in the city is 'satisfied' if he/she lives close to a green area.

The functions plotted show that the offline version marginally outperforms its online counterpart, but only towards the end of the simulation, where the online satisfaction figures decay. This behaviour can be explained after analysis of the pattern of land purchase behaviours in the two approaches. As we will see, key relevant factors in this behaviour turn out to be the spatial positioning of the selected green areas, particularly the distance from the protected cells to the different CBDs, and also the numbers of cells protected.

The Alonso (1964) urban model adopted in this work has a general tendency to concentrate most of the population close to the CBD. Consequently, the measurement of how many of these protected patches of land are close to the three defined central areas of the lattice can provide significant information about the level of satisfaction achieved by the inhabitants of the city. This distance concept is called in the present work the 'closeness' factor. The closeness factor is defined as a measure which averages the distance from the protected cells to each corresponding CBD, see Fig. 6.6. The mean is calculated by grouping the cells of the lattice in concentric annuli around its CBDs. Each ring is considered of distance one. Finally each group is multiplied by its distance and averaged for each time step. This can be formalised as follows:

$$clo = \{\forall c \in P : |\Theta(c) - \Theta(CBD_c)|\} \tag{6.4}$$

where $P$ is the set of protected cells and $\Theta$ is the function that returns the correspondent annuli of a cell $c$ in function of each assigned CBD, denoted by $CBD_c$.

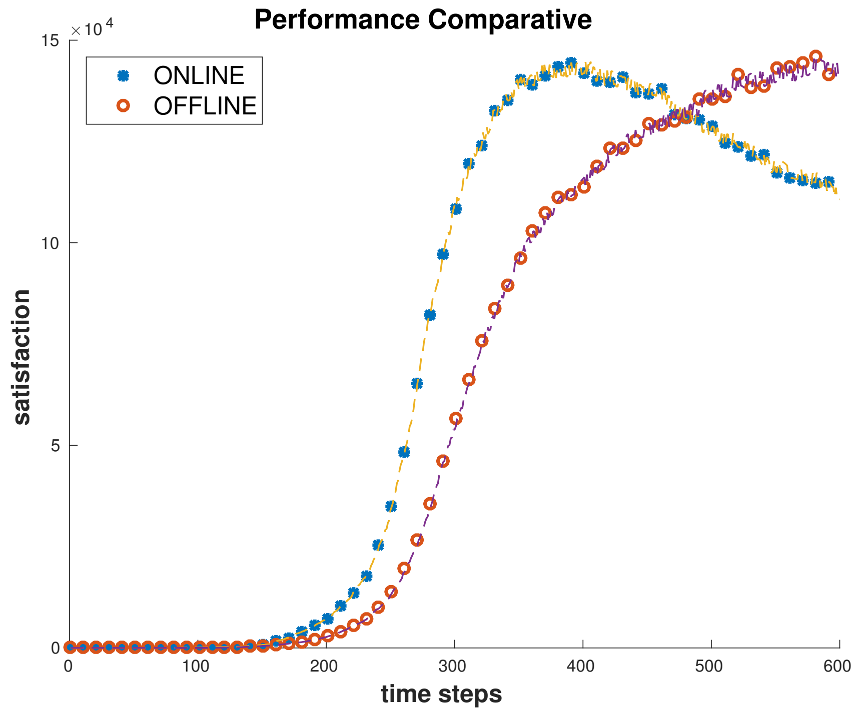



Figure 6.5: Comparing the online & offline algorithms' performance in terms of the satisfaction achieved by the urban population during the complete time horizon of the simulation. This satisfaction is quantified by the application of the fitness function. Regarding the purchasing strategy, the offline approach uses a stochastic strategy to decide the moment in which the purchase decisions are taken and the online approach uses a threshold-based strategy that takes the information from the fitness of the random population of EA solutions as it is explained in Fig. 6.4

.

Here it is important to mention that the difficulty in allocating green parcels of land close to the peri-urban areas of the city (which would improve the closeness factor and better serve more populated areas) is related to the quantitative difference between the available budget and the rural prices of the peri-urban areas of the city. In response to its immediate urban development, the price of these patches of land

| TIME STEP | OFFLINE PERFORMANCE | ONLINE PERFORMANCE |
|---|---|---|
| **50** | 0.00 | 0.00 |
| **100** | 1.44 | 23.64 |
| **150** | 164.26 | 297.58 |
| **200** | 1348.52 | 3433.60 |
| **250** | 6688.92 | 16970.38 |
| **300** | 31299.28 | 72332.92 |
| **350** | 76971.42 | 126669.80 |
| **400** | 107088.92 | 141619.40 |
| **450** | 122133.00 | 140413.40 |
| **500** | 131880.20 | 132985.00 |
| **550** | 138961.60 | 123126.60 |
| **600** | 142662.60 | 115859.60 |

Table 6.2: Numerical values of the online & offline algorithms' performance in terms of the satisfaction achieved by the urban population and measured by the fitness function during the complete time horizon of the simulation (data in line with Fig. 6.5).

increases significantly (Plantinga and Miller, 2001).

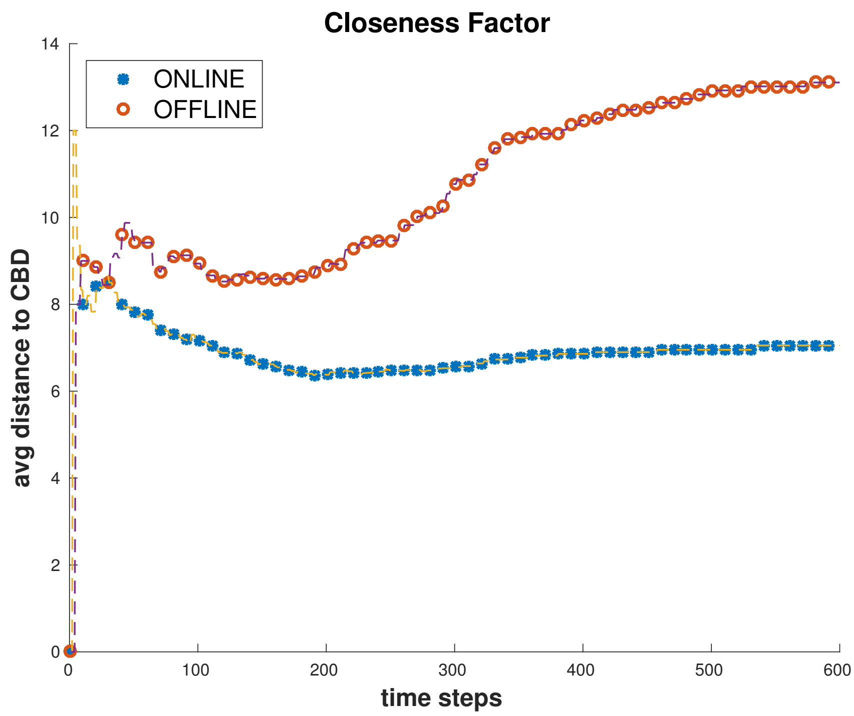



Figure 6.6: This figure represents the average closeness of each protected cell to the corresponding CBD in both algorithms. This factor is calculated using the number of concentric circles between them. Closeness is a key element in the analysis of the performance of both algorithms. This aspect can be seen as a qualitative measure of a given solution.

In terms of the ability of the algorithms to allocate their green cells in more efficient areas of the grid, Fig. 6.6 shows the different values of the *closeness* factor for both approaches. The functions depict noticeable differences in terms of the shape and the average distance from the protected cells to their corresponding CBD. The online approach is a more intelligent strategy, capable of allocating the selected green areas closer to the city centre. This tendency, that can be almost described as constant from time step 200, is valid for the entire simulation period. On the other hand, the offline algorithm displays consistently higher values than the online approach with a steady monotonic increment throughout the process.

| TIME STEPS | ONLINE CLOSENESS | OFFLINE CLOSENESS |
|---|---|---|
| **50** | 8.27 | 8.16 |
| **100** | 9.11 | 7.44 |
| **150** | 8.63 | 6.90 |
| **200** | 8.65 | 6.48 |
| **250** | 9.26 | 6.44 |
| **300** | 10.01 | 6.50 |
| **350** | 11.33 | 6.67 |
| **400** | 11.98 | 6.83 |
| **450** | 12.40 | 6.89 |
| **500** | 12.70 | 6.95 |
| **550** | 12.97 | 6.99 |
| **600** | 13.07 | 7.05 |

Table 6.3: Numerical values of the average closeness of each protected cell to the corresponding CBD in both algorithms, offline and online. Results in line with Fig. 6.6.

Along with the spatial position of the cells, the total number of purchased green areas is also a salient factor which influences the final level of satisfaction. As shown in Fig. 6.7, at the beginning of the simulation both algorithms have very similar behaviour. However from about halfway into the simulation, around time step 350, the online algorithm starts to fall behind in terms of the number of protected cells. This effect is caused by the restrictive purchasing schedule strategy selected for the online approach, which does not consider any affordable non-urban cell profitable enough to be purchased. In contrast, the offline policy constructor is able to more efficiently manage its budget in the latter half of the simulation, buying some affordable cells in the outskirts of the city.

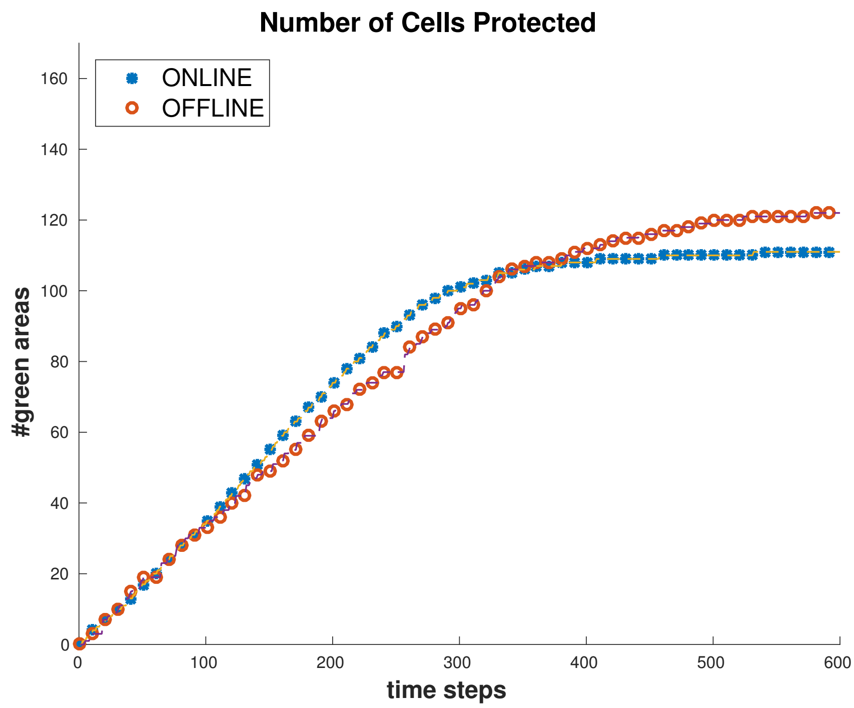



Figure 6.7: This figure shows the performance of both approaches in terms of the number of green cells protected at each time step during the simulation.

| TIME STEPS | OFFLINE NUM CELLS | ONLINE NUM CELLS |
|---|---|---|
| **50** | 8.24 | 8.12 |
| **100** | 25.74 | 25.72 |
| **150** | 41.70 | 44.48 |
| **200** | 57.18 | 64.50 |
| **250** | 72.52 | 82.26 |
| **300** | 86.78 | 96.40 |
| **350** | 101.22 | 103.74 |
| **400** | 108.88 | 107.22 |
| **450** | 114.18 | 108.86 |
| **500** | 117.70 | 109.86 |
| **550** | 120.54 | 110.38 |
| **600** | 121.54 | 111.00 |

Table 6.4: Numerical values of the number of green cells protected at each time step during the simulation by both algorithms: offline and online. These numbers are in line with Fig. 6.7.

One conclusion that can be drawn from these figures is that, if the analysis considers the moment in which the protection purchase was carried out, it can be seen that actions taken at the end of the simulation have less impact on the final performance. This arises from the aggregate nature of the definition of the fitness function. Since the online algorithm needs a certain level of satisfaction to allow a purchase, these late-stage acquisitions are rejected, and consequently we see the offline strategy catching up and overtaking the online strategy in terms of protection decisions late in the city's development.

Future analysis needs to be done to better understand the causes and dynamics of these behaviours. However, it also should be mentioned that the limited number of generations allowed for the online EA may be an important factor. With more generations allowed, for example, it could be that we would see even better protection decisions made at early stages by the online EA, and, although we might still expect the slowdown in purchasing at later stages, the earlier boost may keep it ahead of the offline algorithm in terms of number of areas protected.

### 6.5.3 Spatial Distribution of Cells

Visually, the spatial arrangement of the protected cells resulting from the policy constructor of both approaches, captured in time step 300, shows some noticeably different patterns.

The corresponding figures showing the topological distribution of green areas are the following: Fig. 6.8 represents the final offline spatial distribution with a fixed purchase schedule plan, Fig. 6.9 depict the spread of cells when a stochastic schedule plan is applied and finally Fig. 6.10 that uses the threshold-based strategy to decide when is the best time to buy.

General observations about the distributions of green cells in these three approaches can be made as follows. It is interesting to see how the patterns vary with the different variants of the purchasing strategies. The offline policy constructor implements two strategies: a fixed and a stochastic variant. The threshold-based strategy is developed for the online algorithm. From its qualitative characteristics, the fixed purchasing schedule strategy manages to protect its green areas in locations generally further from the CBDs. This fixed approach also manages to protect a large number of cells, however only a small number of them are placed between the urban cores. Since the approach tries to protect cells as soon as it has funding for any of the available areas and the financial resources received monthly are significantly lower to the prices of the land, to save enough resources to buy expensive cells close to very populated areas is complicated.

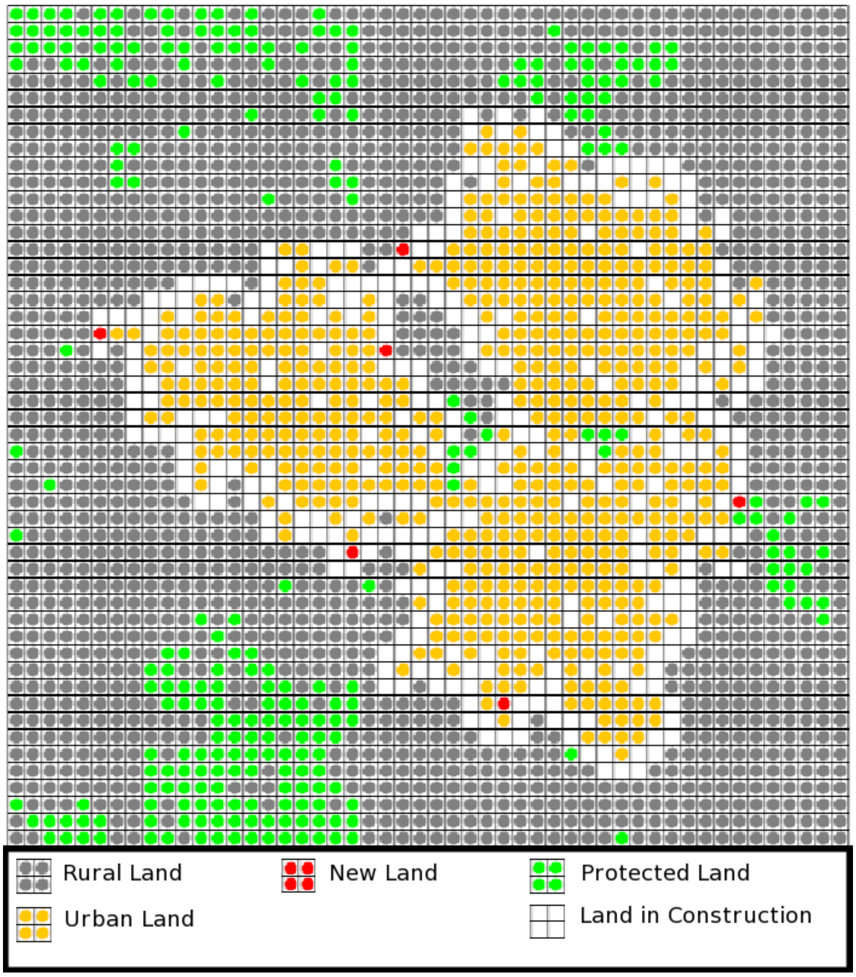



Figure 6.8: Offline algorithm: The lattice shows the spatial distribution of green cells in time step 300 with a **fixed purchasing schedule** strategy. In this case, green areas are located generally further from the CBDs. The approach achieves to protect a large number of cells, however only a small number of them are placed between the urban cores.

In the stochastic purchasing schedule, on the other hand, the total number of protected cells is lower than in the fixed purchasing schedule. Recall that in this strategy the algorithm avoids any purchase 70% of the time that the algorithm has enough resources to do it, which seems responsible for this behaviour. However, the accumulation of money allows the algorithm to invest in more expensive patches of land. Hence, the majority of these cells are located in more populated areas, closer to the CBDs. Then, in conclusion, the stochastic approach surpasses the fixed approach due to the fact that the higher number of cheap protected cells of the fixed approach cannot compete with the higher quality of the ones selected by the stochastic strategy.

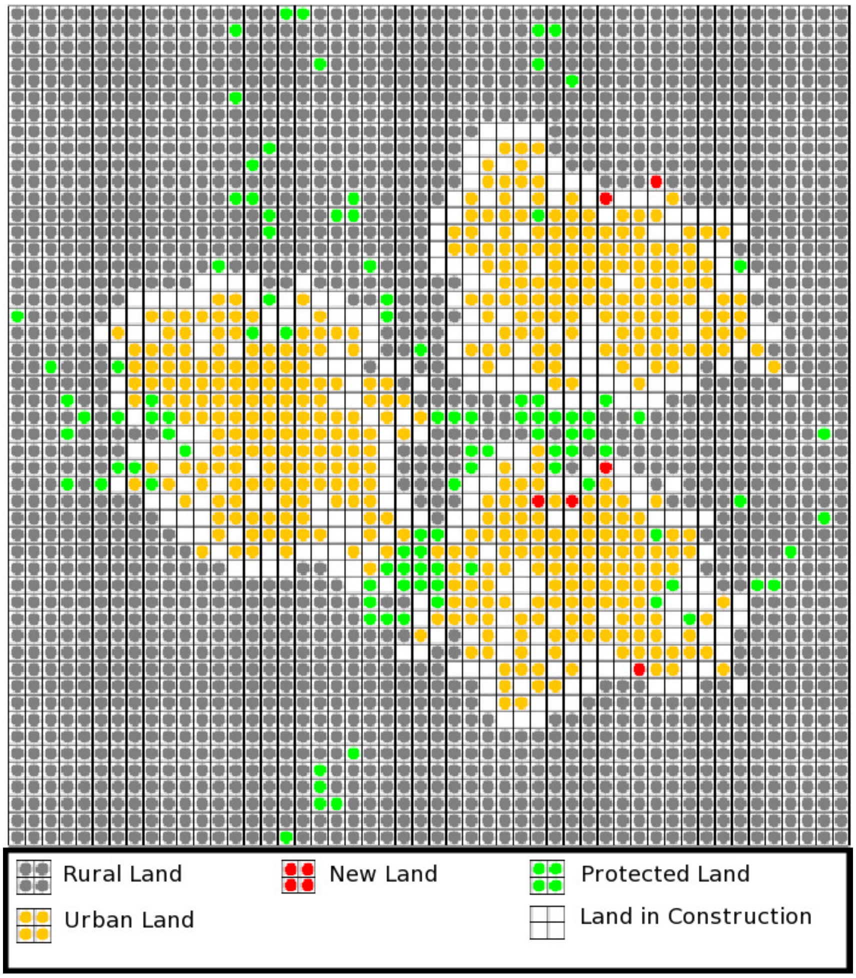



Figure 6.9: Offline algorithm: The grid depicts the spatial distribution of green cells in time step 300 with a **stochastic purchasing schedule** implementation. The total number of protected cells are remarkably reduced if it is compared with Fig. 6.8, a fix purchasing schedule, but the majority of them are located in more populated areas, closer to the CBDs.

The online approach implements a threshold-based purchasing schedule to decide when to buy and when to save the budget for the next generation. This approach is capable of protecting a slightly lower amount of cells than the stochastic procedure of the offline approach at the end of the simulation. It achieves a compact distribution of green areas and connects the majority of them. The quality of these cells also tends to be high, being mostly placed close to the most crowded areas of the city.

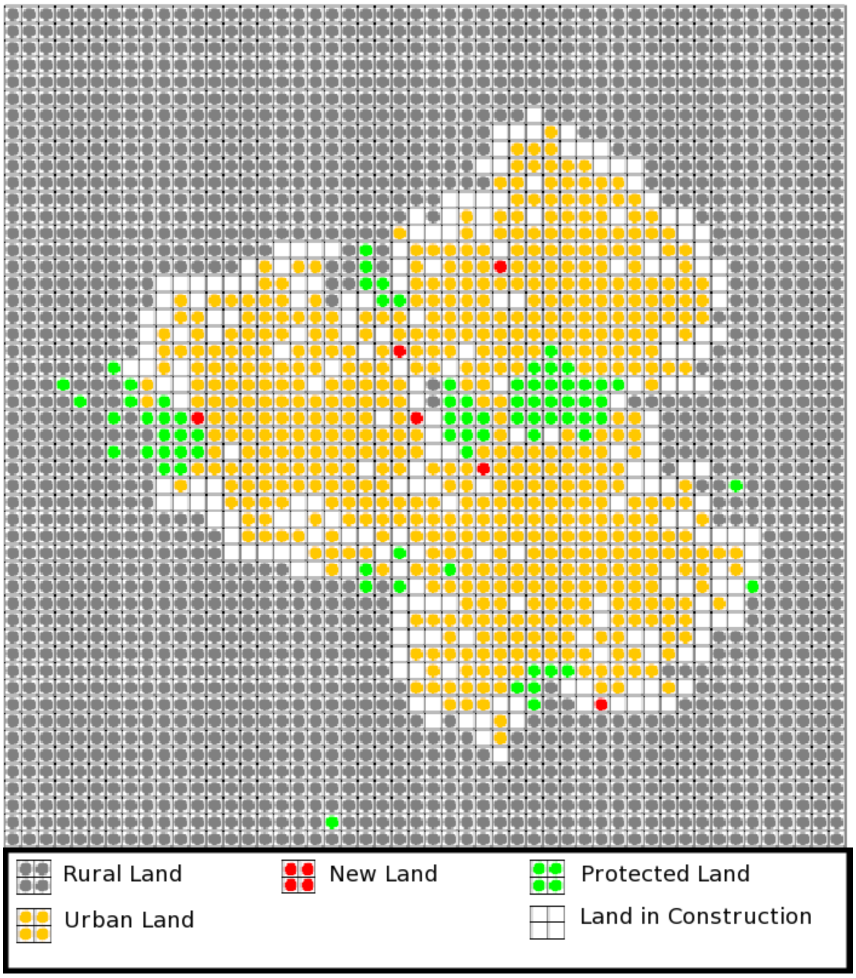



Figure 6.10: Online algorithm: Visualising the spatial distribution of green cells in time step 300 with a threshold-based purchasing schedule implementation. The adaptive strategy achieves the most compact distribution of green areas compared to the other two approaches. With a number of green cells similar to the stochastic strategy, the approach allows to connect the majority of them.

If a comparison between the stochastic offline approach and the threshold-driven online algorithm is carried out, it can be concluded that even if there is a general tendency in both strategies to place more green areas in the zones of confluence amongst CBDs, the offline topological distribution is rather more scattered, with some of the cells located in the extremes of the lattice. In contrast, the offline approach groups the green areas in some well-formed clusters, even though factors like size and compactness of the protected land were not explicitly included as objectives in the definition and calculation of the fitness function.

Focussing now on the offline approach, we can look at the difference between the fixed and stochastic purchasing strategies (Fig. 6.8 and Fig. 6.9). A conclusion that results from inspection of these very distinct patterns is that the purchase strategy seems to be a critical factor for the entire land protection process. The differences (at least based on visual inspection) between fixed and stochastic purchasing appear as marked as the differences between the offline and online strategies themselves. Further consideration allows this to be traced back, again, to the time-dependent nature of the budget, and the strained relationship between budget level and the general dynamics of rural land prices, since the distance to the CBD and the amount of cells are strongly linked with the purchase strategy. Its influence in shaping the overall outcomes leads to the suggestion that ongoing research in this area might usefully consider separating the overall task into two phases, as shown in Fig. 6.11. A key conclusion is the crucial role of the purchase schedule plan, given that purchase decisions are dependent on a restrictive budget which limits the amount of land it is possible to acquire.

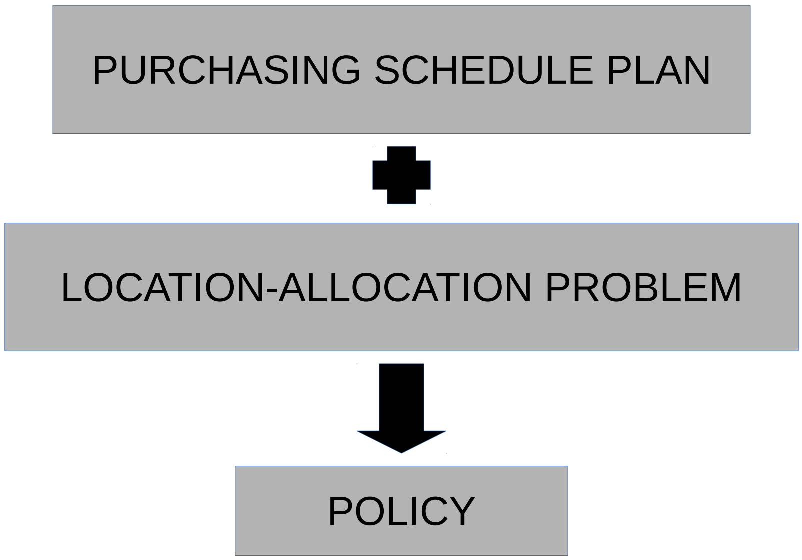



Figure 6.11: Simple depiction of what we propose to be an optimal way to engineer effective green planning practice, based on our experiments and results.

Size is an important factor in the pattern of use of a given green area, linked with the visit frequency and the type of activities undertaken in parks (McCormack et al., 2010). However, even if larger areas are capable of supporting a more diverse type of uses and activities which increases their attractiveness (Broomhall, 1996), other studies conclude that it is better to design a layout with numerous small green areas than a few large parks (Bengochea Morancho, 2003). In this particular aspect, the final spatial pattern distributions linked to the study of the population's needs, as done in this work, clearly produce a contrary conclusion. Even if including other more complex factors into the calculation of the fitness, such as protection areas of high environmental value, crowdedness, level of amenities, size of the park and connectivity with other areas, could affect partially the spatial results, our results can contribute to the discussion towards more effective designs of green area layouts within the green planning community.

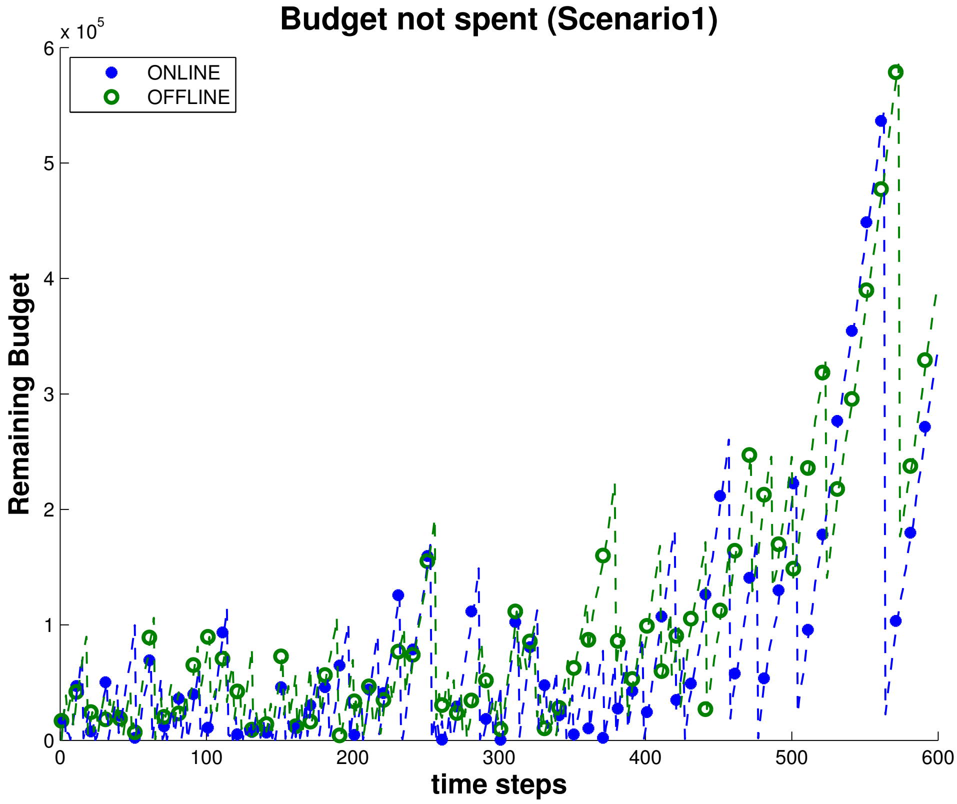



Figure 6.12: In this plot, the behaviour of the budget that was not spent in each turn of the simulation for the online (in blue) & the offline approach (in green) is shown. From time step 400 accumulation of the budget shows the difficulties that both algorithms face to find appropriate cells to buy.

Finally, the behaviour of the remaining budget in one optimisation run using the scenario 1, can be observed in Fig. 6.12. The similar nature of this parameter in both approaches during all the simulation process is an interesting effect to notice. The tendency of the function created can be seen as a measurement of the level of difficulty that both optimisation approaches face in order to acquire new cells to convert them into green spaces. This can be especially observed from around time step 400, where the slope of the functions increase significantly. This is an effect in line with other conclusions held in the system by studying different factors of the optimisation process.

## 6.6 Conclusions

In abstract terms, the field of study of this chapter is SDMPs under uncertainty in stochastic domains. These types of problems are normally addressed with a variety of non-evolutionary strategies. The pursued goal is to assess the suitability of EA techniques for this task, particularly in the context of its application in a green space allocation planning model. The main objective of the developed model is to determine an investment strategy which ensures a proper provision of a given kind of resource, in this case parks, for the population of a city within a predefined period of time.

It is illustrated how two different EA methodologies, an online and an offline approach, are implemented and how they are equipped to cope with the uncertainty in constraints and objective functions, by means of a specific implementation of Monte Carlo sampling. In the online version, the algorithm is able to see the state of the system at any time, being capable of taking decisions in real time. The offline strategy has to look ahead in order to forecast all the possible situations and create an entire plan for a long period of time. Both approaches were tested on a complex version of the urban growth model, configured with three CBDs and multiple price gradients to take more advantage of the application of the evolutionary technique. Finally, the performance of their objective functions, the computational time used and the emerged spatial patterns are compared, explaining the behaviour of both approaches.

The experiments show that both algorithms accomplish similar results in terms of the level of satisfaction achieved by the population, measured by the fitness function, see Fig. 6.5; the offline version marginally out-performs the online at the expense of using significantly more time to converge. Disparity in factors like the number of cells protected, the average closeness to a CBD and the final spatial size and distribution of green areas shows the different planning biases of the algorithms. Due to the aggregated nature of the budget, different purchasing strategies are tested, concluding that the details of the purchasing strategy are critical factors within the problem.

Finally, from the research conducted here we believe that evolutionary techniques can be considered a valid option to solve SDMPs under uncertainty in complex environments.

# Chapter 7

# Conclusion

## 7.1 Introduction

This thesis concentrated on research into applying evolutionary algorithms to a complex problem in urban development. Specifically, the focus of interest was to investigate the creation of plans to locate green spaces over time, subject to budget cycles, in the context of continual growth in both population and infrastructure, the details of which could not be reliably predicted in advance. Particular attention was given to equipping the algorithm with strategies to cope with the inherent uncertainties in the task. Based on Monte Carlo sampling and the use of the urban model as a surrogate, two EA approaches were developed. The first has an offline nature, where the green space allocation policy over the entire time horizon is fully generated at once. In the second 'online' approach, the algorithm solves the problem step by step, and each step is able to adapt with knowledge of the decisions and early outcomes from the previous steps.

## 7.2 Main Findings

The following points are the most important findings reported in this thesis.

- To deal with the uncertainty derived from generating planning policies subject to future unknown characteristics of the system, this work bases the gathering of information on Monte Carlo sampling techniques and on an equivalent version of the urban model used as a surrogate tool. The surrogate model is able to inform the system about the nature of the urban dynamics as well as the amount of noise added to it.

To collect statistics from the surrogate model, the full simulation needs to be run several times to lead to robust results. However, experiments reveal that the number of full simulation runs required for this purpose is not onerous. This conclusion is in line with the analysis of Miller and Goldberg (1996) relating to the optimal sampling for GA.

The statistical knowledge gathered using these techniques is a major factor to take into account for the validity of the results. If the surrogate model is not able to properly infer the trends of the system, conclusions would be poor and non significant. Hence, special consideration in this regard would be required to transfer the methodology to other urban systems.

- This thesis is based on the development of an extension of the traditional urban economic model of Alonso. The model was enhanced to reflect more realistic conditions, in order to fit the requirements of the selected location-allocation problem. These enhancements allow the representation of a mechanism to design policies for acquiring and allocating green spaces in urban areas and to model the ecological degradation process caused by the expansion of the urbanisation.

  The resulting theoretical framework is applied together with a CA-ABM approach, creating a hybrid environment that it is aimed at shedding light on aspects of urban growth. This theoretical model was validated by different experiments conducted and reported in several publications (see Sec. 1.6) and could be the basis of further interesting studies in the area of sustainable urban development.

- This work also shows the applicability of evolutionary algorithm strategies as an appropriate optimisation method for both offline and online versions of the problem. Several aspects of interest from both implementations are:

- One important factor in constructing effective offline planning designs is the adaptive submodularity property that the problem shows under certain configurations of the city. This aspect was analysed using the offline version of the EA and two baselines: one random and another intelligently designed to fulfil the needs of the population. The conclusions derived from the comparison between them are in line with what can be expected when these algorithms are applied to problems characterised as adaptive submodular (Golovin et al., 2011).

  However, even if more complex configurations of the city seem to increase the level of complexity of the search space, favouring the performance of EA over the intelligent CLO baseline, it seems that there are factors that exert more influence than others. In particular, the dynamic nature of rural prices affects the optimisation process more significantly than the addition of other spatial factors, such as creating a more complex spatial configuration of the city with different CBDs. While similar conclusions were reported in other works (Pukkala and Kurttila, 2005), these conclusions are still preliminary and further experiments with more factors should be explored to validate how the complexity affects this type of problems.

- The offline and online method were also compared. In the case of the online method, it can be concluded that the performance of the algorithm increases significantly if, instead of focusing on buying cells as soon as the policy collects enough budget to afford that transaction, the problem is approached from the perspective of finding a policy that sometimes accumulates the current budget in order to facilitate the purchase of a better option in the future.

  To support this, information from the random initial population of the EA approach was used, providing insights into how much the search space is constrained by the budget at this time. This information aided the algorithm in knowing not only when the acquisition operations should be explored, but also to save time in the construction of the online policy. In both cases, due to the complexity of the selected encoding of a problem solution, the crossover operator was removed from the algorithm due to the computational time required in finding feasible solutions after its application.

- The encoding and the selection mechanisms used in the definition of the EAs have a direct impact on the evolution of the system, and hence on the conclusions derived from the experiments included in this thesis. Different encoding would result in other types of case studies with the development of a distinct set of mechanism designed to deal with them.
- Another key aspect of this research is the definition of the fitness function. Its definition is oriented to maintain the level of satisfaction within an urban area which leads to planning solutions which ensure the level of service for this population in the long term. However, due to its concrete formulation, where each agent only gets advantage of the closest green area in the surroundings of its household, the problems falls into the *adaptive submodularity* property commented previously. Similar definitions that account of all the green areas around would have avoided this complication.
- Finally, another crucial aspect for the system is the definition of the budget, with an aggregate behaviour. Similar version of the problem where financial resources cannot be accumulated, would have resulted in a much simple problem. Most of the mechanisms developed in this thesis were aimed to aid the algorithms to achieve a higher level of efficiency due to this factor.

• The results derived from this work could provide some useful insights to the green urban planning community, in particular concerning the spatial arrangement of parks that are designed mainly to satisfy the necessities of the population living close to them. Different spatial patterns were generated according to the different policies that this thesis investigated.

In this regard, the EA tends to naturally cluster the areas close together, which indicates that large green surfaces achieve higher levels of satisfaction from visitors than a set of multiple scattered areas. However, this conclusion is contrary to other results reported in the literature. For example, Bengochea Morancho (2003) claims that it is better to design a layout with numerous small green areas, rather than a few large parks. On the contrary, Broomhall (1996) said that lager areas are capable of supporting more diverse activities, which increase their attractiveness for the visitors. Hence, even if the level of attractiveness of the park was not included explicitly in the model, from the results gathered here it can be concluded that designing large parks are also an option to consider for planners and decision-making stakeholders in order to better fulfil the needs of the people who use them.

## 7.3 Future Work

There are multiple areas presented in this thesis, from both the computational intelligence and computational sustainability perspectives, that would benefit from further work. These possible additional lines of research are discussed in the following subsections.

### 7.3.1 Factors Influencing Agents' Behaviour

As it was previously mentioned, open space planning from a demand perspective (Maruani and Amit-Cohen, 2007), uses attributes from the specific target population to find the most efficient green allocation strategy as a response to social requirements over gardens and parks.

Distance from the park to the dwelling is commonly considered the major factor that influences visit frequency and activities undertaken in parks (Björk et al., 2008; Woolley, 2003). The selection of this feature in planning is based on the observation that only a small percentage of users takes any means of transport to access them (Wong, 2009). Consequently, parks at a short distance are visited more often than large remote parks (Roovers et al., 2002).

With the use of the spatial distribution of the green areas and the population density among other features, a zoning analysis can be used to create a ranking of location alternatives on the basis of their overall attractiveness for the future new park. The process includes conducting a multicriteria land suitability analysis, and the posterior selection of optimal sites through an optimisation process according to a measurable criterion.

However, even if from this quantitative perspective the provision of green areas has been carried out following an efficient process, this fact does not necessarily imply that users have enough incentive to visit them. In this matter, a very small number of studies are focused on the factors that could promote or limit the use of these green areas (Bixler and Floyd, 1997; Hitchings, 2013).

In the concrete case of green space allocation, the existing literature covers a specific range of issues, such as the protection and restoration of valuable and degraded areas (Zucca et al., 2008), the preservation of carbon stocks (Marinoni et al., 2009) or the definition of ecological corridors (Ferretti and Pomarico, 2013) among others. In these studies, data are normally gathered by on-site surveys, a series of spatial observations and by experts' knowledge. In the present proposition of future work, we are especially interested in the analysis of factors which negatively influence the frequency of visits to a park.

According to that premise, future studies could investigate some of these factors, such as crowdedness, size, distribution and design, in an urban growth model using an Agent-Based System framework in order to improve the understanding of the individual perceptions that directly influence the frequency of use to these open spaces. This knowledge can contribute to the design of more comprehensive green policies which enhance the satisfaction of a larger number of residents. Based on these factors, the model, the behaviour of agents and the nature of the fitness function can be enriched to better mimic heterogeneous behaviours of visitors.

In the following, we discuss in particular the factors that could discourage people to visit parks:

#### 7.3.1.1 Crowdedness

High population density in green areas is associated with various factors including population and urban growth. A higher demand for these spaces is created by the rise of the average standard of living, and an increasing level of environmental awareness in society (Cheshire and Sheppard, 1998; Kline, 2006; McPherson, 2006).

Crowding perception is a subjective concept which can be perceived when the area is highly congested with heavy pedestrian and vehicular use. Depending on the specific user profile, the perception of crowdedness may be different. For instance, fairly experienced users feel more intensively the saturation when they compare their current visit with past experiences (Ditton and Sutton, 2004; Vaske et al., 1980) and when they use the area more frequently (Arnberger and Brandenburg, 2007). These local visitors can feel the saturation to be a factor which decreases their quality of life (Lankford and Howard, 1994; Brunt and Courtney, 1999; Williams and Lawson, 2001).

Congestion can provoke different compensatory measures such as time and intraspatial displacement if any suitable alternative exists (Hall and Shelby, 2000; Shelby et al., 1988; Manning and Valliere, 2001), and may decrease the importance of the distance to the park for its use (Kaczynski et al., 2008). This interspatial displacement also implies extra costs in terms of time and transportation, which can be a problem for low income individuals who cannot afford to move to other green areas outside the city (Scottish Natural Heritage, 2008).

#### 7.3.1.2 Size & Distribution

Size is an important factor to take into account when patterns of use of green areas are analysed. In the related literature, it is a common practise to group parks into two different types: urban local green areas selected for daily outdoor activities, and non-local areas used for excursions or weekend sports (Arnberger, 2006). McCormack et al. (2010) consider that larger parks are capable of supporting a wider range of activities, which increases their attractiveness (Broomhall, 1996). From a different perspective, Bengochea Morancho (2003) concludes that it is better to have numerous small green areas, which should be complemented with a well planned set of large parks. In her analysis of how green areas influence prices of the households, this conclusion is derived from the role of distance, which is primarily emphasised over others.

There are studies that connect the concept of distance with size. Pouta and Heikkilä (1998) create a classification relating type of green area, size and distance. They define the minimum size for local parks to be between 1.5 and 3 hectares reachable in 300 metres, and outdoor recreational parks are characterised by 20-25 hectares at 1 kilometre. The European Commission has recommended that residential proximity to green spaces should be limited to 300 metres with an area of at least 5000 $m^2$ (Tarzia, 2003).

Another important aspect to consider is the topological distribution of these areas within the city. It is quite common find cases where parks are non-homogeneously distributed. Instead, they are generally concentrated over some districts which leads to extensive areas with a lack of proper provision.

A non-homogeneous set of green areas contributes to the appearance of inequalities where some people have easier access to nature areas in their local neighbourhoods than others (Pickett et al., 2001). In general, underprovision and an overall lower level of vegetation cover are more commonly found in low income areas (Iverson and Cook, 2000; Pham et al., 2012). This factor is an important environmental equity issue for city planners.

#### 7.3.1.3 Design & Green Services

Design is another important element which influences the frequency of use of green areas (Schroeder and Daniel, 1982) and contributes to the improvement of health and well being (Floyd et al., 2008). Users travel further distances to visit a certain green area if it has extended characteristics and enhanced aesthetical factors (McCormack et al., 2006; Epstein et al., 2000). Commonly size and design are also concepts linked together since larger parks permit the allocation of a wider offer of services (Giles-Corti et al., 2005).

However, as a negative factor, the level of greenery and physical barriers like inadequate facilities to interact with the park (walking trails), lack of transport choice, poor accessibility or unaffordable recreational activities may discourage some people to use these parks.

### 7.3.2 Enrich Population profiles in their use of Green areas

Population segments divided by gender, age, household composition and socio-economic status differ in how they use and perceive green areas (Burke et al., 2009; Eisler et al., 2003). This makes it very challenging to find a unified policy which achieves a complete fulfilment of these diversified demands. For instance, it can be mentioned that elderly people show lower frequency of use due to personal mobility, health and security fears (Payne et al., 2002; Burgess et al., 1988). Meanwhile children have higher needs of open areas for playing and social interaction when they live in high populated dwellings (Loukaitou-Sideris and Stieglitz, 2002; Crane et al., 2006).

It is also proposed to make a more active use of population demographics and background information for grouping visitors according to different interests and personal characteristics using an ABM approach. As it has been argued, ABM is a technique particularly suitable for studying socio-economic and environmental trends based on heterogeneous individual interactions and it complements other equation-based techniques by means of the exploration of individual-level behaviours (Brown et al., 2006). ABM has been extensively used to study urban growth phenomena (Huang et al., 2013; Matthews et al., 2007) in the context of rich individual profiles, normally parametrised from quantitative surveys (Robinson et al., 2012).

### 7.3.3 Multi-objective Approaches

Since problems related to location-allocation of resources are in nature multiobjective (Watts et al., 2009; Nelson et al., 2009), we propose the implementation of a multiobjective planning extension of our current urban model (Vallejo et al., 2013), focused on the discouraging factors commented on the previous section: crowdedness, low level of activities, poor accessibility, lack of security. The analysis of these elements in a diversified population allows us to capture and understand the most relevant synergies and conflicts created by the interactions with other dynamics included in the model, such as demographic growth, urban extension and environmental degradation.

#### 7.3.3.1 Ecological Protection

One of the most urgent research issues within the broad field of urban planning is the study of mechanisms that can mitigate the ecological degradation that is invariably linked with urban expansion. What makes this particularly difficult is that the process of urban expansion needs to achieve effective and acceptable results at many time-scales. For example, if a growing city builds quickly on the majority of the green spaces available to it, it will severely limit its further growth opportunities. An optimal land-use allocation can preserve valuable land resources and maintain environmental stability.

Another important kind of landscape that can be located in the surroundings of urban areas is forest landscape. Urban forestry refers to concentrated groups of trees in urban settings. The use of these landscapes has been oriented according to a set of different goals, often contrary to each other (Nilsson and Randrup, 1997), such as timber production as a provision service in contrast to the ecological interest in maintaining natural diversity and boosting regeneration capacity. This can also facilitate the reduction of $CO_2$ emissions from deforestation and soil degradation.

To manage such conflicts, it is necessary to achieve a trade-off between the different aspects, so that the urban forestry program emerges as a state funded entity to manage a sustainable production of timber. The main agents involve in this process are:

- State forest service.
- Private owners.
- Wood processing industry.

- Environmental NGOs.

This line of research would also focus on public ownership as a tool to fulfil the conservationist goal. Municipalities would acquire these areas using what is called 'land-buying funds', and fitness would measure not only the amount of protected cells in the lattice in terms of the ecological value at the time of acquisition, but also the number of other cells that are protected in its neighbourhood. This way, the fitness measure guides the process towards promoting larger, connected protected areas, which is a crucial issue in terms of sustainability.

Formally this fitness function could be expressed as follows:

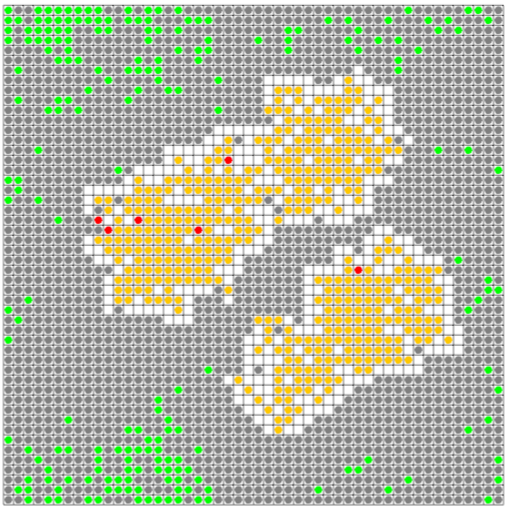
$$f(t) = \sum_{i=0}^{T} (bio(c_i) * 100 + NV \times Ncp(c_i)) \tag{7.1}$$

where $c(t) \in P$ is the set of protected protected cells in time $t$, $bio$ is the function that retrieves the ecological value of the cell $c_i$, $NV$ is a constant that represents the additional reward the fitness has for each protected cell that belongs to the neighbour of the cell $c_i$ into consideration and $Ncp$ is the function that returns the number of protected cells in the neighbourhood of $c_i$. $bioValues$ are in the range $(0-1)$ . To calibrate the effect of this factor in the calculation of the fitness function, it is required to increase its magnitude by a factor of 2.

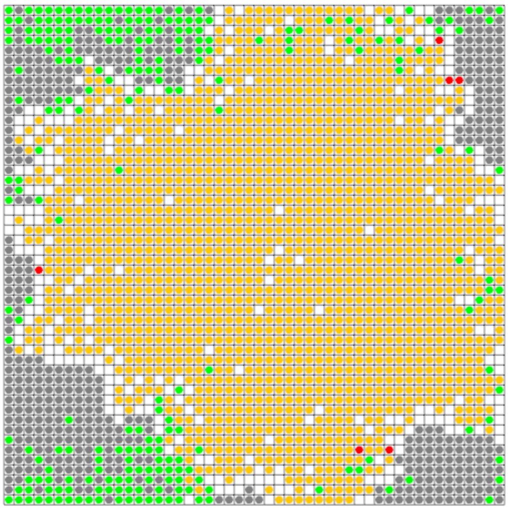
The proposed future work would be focused on the optimisation of the two conflicting objectives: (i) to achieve the maximum level of service for the population living in the city, and (ii)maintaining a high level of ecological protection where these values take into account as a positive factor the amount of green areas located together, which can consequently support better a large number of biota. The problem can be represented with the following vector $J$ of $n$ system responses, where $n$ in this case is equal to 2.

$$M = \begin{bmatrix} J_1 \\ J_2 \end{bmatrix} = \begin{bmatrix} Satisfaction \\ Protection \end{bmatrix} \tag{7.2}$$

In continued but uncompleted work in this direction, we have applied a modification of the PAES algorithm as the MOEA approach to find the best trade-off between these goals. The spatial representation of the implementation of this algorithm using the satisfaction fitness and the conservation fitness leads to arrangement of green areas below.



(a) Visual representation of the city and the generated policy at time step=300.



(b) Illustration of the final arrangement of the city at time step=600.

These figures shows that, even if it is possible to generate a policy that covers both conflicting aspects, the policy constructor seems to favour areas further from the centre of the city due likely to the very affordable nature of the prices in these areas. More research and further analysis would be necessary to come up with further conclusions.

### 7.3.4 Model extension

The proposed extensions of the urban model can be outlined as follows:

- Defining different agents' profiles according to their family structure, gender and age factors. Each profile should have associated a set of predefined motivations to visit green areas, which generate different patterns of use. These varied types would be based on data collected from quantitative studies published in the associated literature and could include: elderly people, young families with children, teenagers, young professionals and so on. Since age and offspring are characteristics that have been already implemented in the model, the task here will consist only in the mapping of each of them to a set of predefined profiles, adding the corresponding information to each agent. The profile will cause some effects on the computation of the fitness, contributing to each profile with a different level of satisfaction according to the type of green area in the surroundings. From a computational point of view, this will not affect the performance of the model. However, since population is used to compute the fitness, richer information would be necessary to be collected to aid the EA algorithm to construct an efficient policy. Further checking would be required to see if this information could be easily collected and transferred to the algorithm.

- Defining the rules that allow the system to decide for each agent which parks would be visited and with which frequency. Taken into consideration are their location in the grid and their patterns of use according to their personal profile. Using the resulting population density and the size of each park, this line of work would include the concept of crowdedness, which will decrease the level of satisfaction of the affected population. This assignation could add some complexity to the system if the agent is able to choose an alternative green area in their surrounding, is this area exists or if the use of this areas can be shared between some of them. The potential of this type of model is in the study of complex dynamics of use of the population. This can be extended to include the use additional means of transport to rich other green areas that offer further level of amenities.

- Clustering analysis. Based on the preliminary visual analysis of the topological structures formed by the spread of green areas generated by the different algorithms developed in this thesis, further quantification is needed to draw consistent conclusions about the consequences for the population. The study could be implemented using multiple metrics, like size, shape and relative location of these areas, and techniques such as k-mean algorithms based on the topological centroids. This aspect can be extended to study the links between green areas, applying techniques like connectivity models based on distance, like hierarchical clustering. This extension would require extra computation but from a post-processing perspective. Another module within the framework could be created to take as an input the final arrangement of the cells in the lattice and compute connexions, dependencies and relationships between these green areas.

- Enriching the characterisation of green parcels to include the aforementioned factors and their interdependences (design, green services, accessibility). Adjacent green areas will be considered as a single unit. This will permit the existence of parks larger than a single cell in size with different designs, levels of greenery and activities to be undertaken. To implement this characteristics within the model a new class that abstracts these green areas over the single unit green areas will be implemented. The division of functionalities will be required to split the code between them and create appropriate methods of communication. Other elements should be also revisited like distance to green areas, which should look for the centroid of the park and the ecological value of the area, which could not be homogeneous.

- To model the crowdedness factor each park should have a metric related to the potential number of visitors than can supply according to their size. This metric would be constructed in function of the people living close by. Further refinements would be needed to capture that a given agent can visit different parks if multiple areas are located in their surroundings. This selection would be affected by multiple factors including the level of amenities, the profile of each particular agent and the accessibility level. Crowdedness will also influence the fitness function as a negative factor. However, the perception of crowds are subjective. An heterogeneous influence should be added. This should be extended to the variety of agents' profiles, defining different upper and lower bounds for each of them. Further analysis should be carried out to measure this characteristic as a general aspect of the area.

- Creating some scenarios in which the final spread of parks are not homogeneously distributed, in order to study how the population could cope with crowded green areas or a lack of provision. The use of scenarios has proven useful in the study of the impact of multiple socio-economic dynamics in an urban context (Murray-Rust et al., 2013). In each scenario, the model could be optimised using an EA technique for a single or multi-objective set of characteristics, encompassing the minimisation of economical costs and the maximisation of both the population satisfaction and the protection of the highest ecological valued areas. In this case, instead of starting the study from an empty grid, different more mature urban configurations would be the starting point. Extra functionality would be required to allow these types of lattices and extra gathering of data would be implemented to study the consequences of each of these scenarios.

- The study of the time required to run the experiments will be extended to cover other areas of the thesis. Information should be added regarding the length of the experiments and how this factor changes when different configurations of the city are tested. This will cover multiple sizes of the lattice and richer layouts, with more CBDs located differently within the grid. Output information from the simulation would include this extra information, with no extra additions to the model or optimisation procedures.

# Appendix A

# Data & Coding

This appendix includes information regarding the code and experiments from which this thesis is underpinned. The entire content is stored in a public repository in the GitHub community, concretely in `https://github.com/MartaVallejo/PhD_Code`. The code is stored under the GNU General Public License v3.0 and can be cited with DOI:10.5281/zenodo.802116.

The structure of the stored information is the following:

- Code: stores all the modules that the urban model framework is compounded. All the files are classes implemented in Java.
- Experiments: allocate the data gathered from the different experiments conducted during the PhD. Each experiment has a different name scenario. Concretely, there are four scenarios:
  - Scenario 6d
  - Scenario 9gc
  - Scenario 10b
  - Scenario 13
- Matlab Scripts: covers all the scripts created to post-process the data generated by the model.

- *developmentNotes.pdf*: file with information about how to run the experiments, descriptions of the modules, main variables, input and output files and Matlab scripts.

# Appendix B

# Development Notes

## B.1 General Description

### B.1.1 Modules & Files

- Random: Random approach, non-optimised.
  - Baselines.txt
  - SatisfactionRan.txt
- GatherData: Used only to gather the statistical data.
  - Density.txt
  - NonUrbanPrices.txt
  - Urbanised.txt
  - Rings.txt
- ClosestHeuristic: The closes cell that we can buy (from Scenario 7)
  - Baselines.txt
  - SatisfactionClo.txt
- Genetic algorithm: GA_Satisfaction and GA_Distance.
  - GA_DATA.txt General info

  - Result_GA.txt with all the cells protected

- TestOptimisation: Test the EA approach.
  - SatisfactionGA.txt
  - Inconsistencies.txt
- MultiOptimisation: online GA Satisfaction (from Scenario 13)
  - SatisfactionMO.txt
  - realfitness.txt
- CheckFiles:
  - Check Visually the distribution of cells protected in *Result_GA.txt* or *Satisfaction CLO, RAN*
  - Check *Result_GA.txt* regarding the budget.
  - Visualise the population distribution in a given tick.

### B.1.2 Variables

Seven kind of prices, each with:

1. Fitness:
   - Distance **only GA**
     - $Tick_0 - Min(Tick, sat) \neq 0$
     - $Min(Tick, sat) \neq 0$
   - Satisfaction
     (a) Ticks counted **only GA**
       - Accumulative. $\sum_{i=1}^{S} \sum_{j=1}^{C} \sum_{k=t_j}^{N} s(c_{i,j})_k$
       - Single.
         * Initial $\sum_{i=1}^{S} \sum_{j=1}^{C} s(c_i)_{t_0}$
         * End $\sum_{i=1}^{S} \sum_{j=1}^{C} s(c_i)_{t_N}$

(b) Per agent

- Unique: Maximum satisfaction per agent: $\forall a \in A, s(a) = \max s(a,l)$
- Aggregated: Total satisfaction per agent: $\forall a \in A, s(a) = \sum s(a,l)$
  * Flat: only one point per park if it is at maximum 3 cells of distance.
  * Weighted: (1 to 3) point depending of the distance.

2. Selection creation process: **only GA**

- Fix: all possible cells.
- Stochastic: only a percentage of possible cells.

### B.1.3 Files

Data gathered for 600 ticks of the clock.

#### B.1.3.1 Gather Data

For each time step, we store an entire lattice $50 \times 50$ cells (2500). The lattice is made up by a set of integer values. Each cell store a value in the system.

- **Non urban prices.**
  - The number of times that this row has been updated.
  - Mean of the price of the land when city grows.
- **Urbanised**
  - Number of times this row has been updated.
  - If the cell has been urbanised to the time t, then sum one in the lattice correspondent to this time step.

$$1 + SizeLattice$$

- **Density**

  - The number of times that this row has been updated.
  - If the cell is urbanised then it is calculated and stored the mean of the previous value and the new value. Change: only sum and it can be divided later.

- **Rings** Non-urban averaged priced in the different rings of the lattice.

#### B.1.3.2 TestSolution, Random, Closest Heuristic & Multioptimisation

*TYPE_FITNESS*: S(Satisfaction), D(Distance).

*TYPE_SELECTION*: F(Fix), S(Stochastic).

*TYPE_SATISFACTION_AGENTS*: Flat: 1 / Weighted: 2

*TYPE_FITNESS_TICKS*: Accumulative (10), Single - Initial: (21) / End: (22)

*TYPE_OPTIMISATION*: 1, 0, MixApproach (TODO) In Multioptimisation

*SAVE_DENSITY*: 1 Save population distribution to calculate real fitness.

- **Satisfaction** Different runs are appended into the same file if the run is not finished properly. Each line corresponds to one time step.

  1. Total population.
  2. Total satisfaction (1): All the green spaces count.
  3. Total satisfaction (2): Only the closest green space counts.
  4. Total satisfaction (3): Protection
  5. Total urban cells.
  6. Total protected cells.
  7. Min Green price.
  8. Max Green price.
  9. Average Green price
  10. Min Urban price.
  11. Max Urban price.
  12. Average Urban price

13. Min populated
14. Max populated
15. Migration
16. Closeness to the CBD
17. Protected cells
18. Remaining budget

**SatisfactionMO (Additional fields):**

19. Time
20. GA Population size
21. Generations until convergence
22. Number of mutations
23. Worst fitness
24. Best fitness
25. Mean fitness
26. SD fitness
27. 0 (not evolved), 1 (evolved not used), 2 (evolved and used)

**In SatisfactionGA (name file):**

1. GA ID (4 digits).
2. Satisfaction ID (4 digits).
3. Type of fitness: (*TYPE_FITNESS*)
   (a) Satisfaction (S) - Both aspects
      – Ticks (*TYPE_FITNESS_TICKS*)
         * Accumulative (10)
         * Single - Initial: (21) / End: (22)
      – Agents - Flat: 1 / Weighted: 2 (*TYPE_SATISFACTION_AGENTS*)

   (b) Distance (D)

4. Type of selection. (*TYPE_SELECTION*)
   (a) Fix (F)
   (b) Stochastic (S) - <percentage> (*PERCENTAGE_STOCHASTIC*)
5. Feasibility
   (a) Feasible (F)
   (b) Infeasible (I)

Example: 1234 5678 S101 F I - 1234 5678 D S50 F

**In SatisfactionMO (name file):**

1. GA ID (4 digits).
2. Satisfaction ID (4 digits).
3. Type of fitness: S (*TYPE_FITNESS*)
   - Ticks (*TYPE_FITNESS_TICKS*)
     * Accumulative (10)
     * Single - Initial: (21) / End: (22)
   - Agents - Flat: 1 / Weighted: 2 (*TYPE_SATISFACTION_AGENTS*)
4. Type of selection. (*TYPE_SELECTION*)
   (a) Fix (F)
   (b) Stochastic (S) - <percentage> (*PERCENTAGE_STOCHASTIC*)

Example: 1234 5678 S101 F - 1234 5678 S101 S50

- **Budget** Budget used to buy and protect land. One per time step. Field: Amount of budget given to the municipality in each time step. It is generated randomly.
  - BUDGET_R: Randomly generated
  - BUDGET_P: In function of the population
- **Scenario** Initial biovalues of the cells included into the lattice. Type: double.

#### B.1.3.3 GA Phase

- **ResultGA** Cells selected to be protected according to the GA optimisation procedure.

  File content:

  1. Tick of the clock only when a cell is protected: it can be repeated if two cells are protected in the same slot.
  2. Pair/s of coordinates of each cell.
  3. Price of the cell.

  Name Format:

  - Times that statistics are gathered.
  - Type of fitness: (*GA_Satisfaction* or *GA_Distance projects*)
    1. Satisfaction (S) - Both aspects
       * Ticks (*TYPE_FITNESS_TICKS*)
         · Accumulative (10)
         · Single - Initial: (21) / End: (22)
       * Agents - Flat: 1 / Weighted: 2 (*TYPE_FITNESS_AGENTS*)
    2. Distance (D)
  - Type of selection.
    1. Fix (F)
    2. Stochastic (S) - <percentage>
  - Feasibility
    1. Feasible (F)
    2. Infeasible (I)
  - ID (Four digits).

  Example: 20 S101 F F 1234, 20 S212 S50 I 1234, 20 D F I 1234, 20 D S50 F 1234

- **GA_DATA** Located in folder *Workspace* to be updated from both GAs.

  1. Type optimisation: OFF_SAT, ON_SAT
  2. Lattice size
  3. GA Population size.
  4. Ticks of the simulation.
  5. Times that statistics are gathered.
  6. Generations until convergence.
  7. Number of mutations.
  8. Worst fitness.
  9. Best fitness.
  10. Mean fitness
  11. Standard Deviation SD
  12. ID (Same than Result_GA).
  13. Type of fitness:
      (a) Satisfaction (S) - Both aspects
        - Ticks
          * Accumulative (1) - **0**
          * Single (2) - Initial: 1 / End: 2
        - Agents - Flat: 1 / Weighted: 2

      (b) Distance (D) - **0 - 0 - 0**
  14. Type of selection.
      (a) Fix (F) - **0**
      (b) Stochastic (S) - <percentage>
  15. I/F Infeasible solutions - Feasible solutions
  16. Scenario (Type of prices).
  17. CBDs.

18. Type of Budget

19. Computational Time

Example:15 600 20 xxxx xxxx xxx xxx 1234 S 1 0 1 F 0 9g, ...1234 D 1 0 1 S 50 9F

Mean and SD fitness should be calculated in MATLAB

- GA Parameters

  1. Number of individuals in the population.
  2. Number of simulations.
  3. Size of the lattice.
  4. Scenario.
  5. CBDs info.
  6. Feasible - infeasible solutions.

#### B.1.3.4 TestOptimisation

- **Inconsistency** One file for all runs.

  1. GA ID (4 digits).
  2. Satisfaction ID (4 digits).
  3. Number of inconsistencies: times that we try to protect a cell that is already urbanised.
  4. Number of failures because of a lack of budget.
  5. Times statistics are collected.
  6. Type of fitness:
     (a) Satisfaction (S) - Both aspects
        - Ticks
          * Accumulative (1) - **0**
          * Single (2) - Initial: 1 / End: 2

- Agents - Flat: 1 / Weighted: 2

(b) Distance (D) - **0 - 0 - 0**

7. Type of selection.

(a) Fix (F) - **0**

(b) Stochastic (S) - <percentage>

8. Feasibility

(a) Feasible (F)

(b) Infeasible (I)

9. Scenario.

10. CBDs.

11. Size lattice

12. Total Time

13. Type Price

14. Type Budget

Example: 1234 5678 1 15 20 S 1 0 1 F 0 9g, xxx D 0 0 0 S 50 9f

#### B.1.3.5 MOOptimisation

- **Inconsistency** One file for all runs.

1. GA ID (4 digits).

2. Satisfaction ID (4 digits).

3. Type of fitness:

(a) Satisfaction (S) - Both aspects

- Ticks
    - Accumulative (1) - **0**
    - Single (2) - Initial: 1 / End: 2
- Agents - Flat: 1 / Weighted: 2

   (b) Distance (D) - **0 - 0 - 0**

4. Type of selection.

   (a) Fix (F) - **0**

   (b) Stochastic (S) - <percentage>

5. Scenario.

6. CBDs.

7. Size lattice

8. Time

Example: 1234 5678 1 15 20 S 1 0 1 F 0 9g, xxx D 0 0 0 S 50 9f

- **Baselines**

  1. CLO/RAN
  2. id
  3. Scenario name
  4. CBDs
  5. Size lattice
  6. Budget
  7. Total time

### B.1.4 Scenarios

#### B.1.4.1 Backup Allocation

At University: *mv59/private/versions*

Home: *Documents/PhD/CollectedData/Code*

## B.2 How to run experiments: Steps

All the programs should share the same *scenario.txt* file and *budget.txt*. Create a folder with the name of the new scenario if it is necessary and place the files into the folder: CollectedData/Calculos.

- In the GatherData module.
  - Check current size of the lattice and change it if it is appropriate.
  - If we want to start from scratch, initialise Urbanised, NonUrbanPrices & Density files to zero.
    * Location of the files: collectedData/Scenarios/CaseBase/FilesToZero.
    * Take the ones with the correct size of the lattice and with the proper number of ticks of the clock.

    Otherwise copy them form the case we want to continue.
  - In *lattice.java*
    * update *CBDS* according to the desired scenario. *NUM_CBDS* will be updated automatically
    * update scenario to assign the proper non-urban prices in *cell.java*.
  - Run *nLattice::TOTAL_TICKS* times GatherData to collect statistical data for the GA in the Urbanised, NonUrbanPrices, Density & Rings files.
    * Create a folder with the time that the data is gathered.
- In the GA Satisfaction/Distance project (normal Java project with parameters).
  - Check input parameters: population size, number of generations, size lattice and scenario name.
  - Update Density, NonUrbanPrices & Urbanised files with the new files calculated before.
  - Run the program.
  - Copy *Result_GA.txt* to CollectedData/Scenario/NomScenario/Num_Runs when the program finishes
- The file *Workspace/GA_Data.txt* is automatically updated.
- In TestOptimisation project:

  - Parameters: size of lattice: in *<projectname>.rs context.xml*::"World Size", "width", "height" (source view).
  - In *lattice.java* update:
    * Number of runs *(TOTAL_TICKS)*
    * *NUM_CBDS* and *CBDS* according to the desired scenario.
    * scenario to assign the proper non-urban prices in *cell.java.*
    * *SCENARIO_NAME* with the name given to the current scenario.
  - Copy the *Result_GA.txt* file from GA that we want to check.
  - Run the program
  - Copy *satisfaction.txt* into the folder CollectedData/Scenarios/NomScenario/Num_Run.
  - The file *inconsistecy.txt* is automatically updated.

- When the collection of data is finished for the current scenario, copy *GA_Data.txt* and *inconsistecy.txt* to CollectedData/Scenario/NomScenario
- In Random project & Closest project
  - Check parameters: size of lattice
  - In *lattice.java* update:
    * number of runs *(Lattice::TOTAL_TICKS)*
    * *NUM_CBDS* and *CBDS* according to the desired scenario.
    * scenario to assign the proper non-urban prices in *cell.java.*
  - Run Random / Closest
  - Place & *satisfactionRAN.txt* into CollectedData/Scenarios/NomScenario/RAN or *satisfactionCLO.txt* into CollectedData/Scenarios/NomScenario/CLO as corresponding
  - *protected.txt* is not currently used.

If there are enough changes to create a new scenario, copy the *.java* files into folders, zip them and copy to the folder *CollectedData/Code*

### B.2.1 List of Functions in Matlab

To run the post-process scripts in MATLAB, go to *myFunctions2.m*, choose the number of the function to run, see if *matlabXX* needs any adjustment and call *myFunctions2(num_function)*.

1. **budgetGenerator** Generate a file budget dependant on the population alpha = 0.5: measure the importance of the population when the budget is generated. *Density.tx*t: File with population evolution

2. **simpleSatisfaction** (matlab) Show simple satisfaction comparative for EA, MO, CLO and RAN

3. **simpleSatisfactionBar** (matlab) Show simple satisfaction comparative for EA, CLO and RAN

4. **satisfactionTable50** (matlab) Show simple satisfaction comparative for EA and RAN

5. **latexTable** (matlab2) Create the code in latex for a table with the satisfaction achieved by the three approaches. Changes in the folder of the script *matlab2* should be done to gather data for the three scenarios.

6. **bestSatisfactionTest** (matlab) *Inconsistency.txt* Create a plot with the best Test values achieved with one inconsistency for each scenario.

7. **failuresPlot** (matlab) *Inconsistencies.txt* Create a stacked bar chart using the bar function with information about the number of failures due to budget and due to city expansion.

8. **populationPlot** (matlab) Create a plot with the population behaviour

9. **cellsUrbanisedPlot** (matlab) Create a plot with the number of cells urbanised for RAN, EA and CLO.

10. **protectedCellsPlot** (matlab) Create a plot with the number of cells protected

11. **migrationPlot** (matlab) Create a plot with the behaviour of the migration

12. **lowerGreenPricesPlot** (matlab) Create a plot with the collected green prices with the lowest value

13. **higherGreenPricesPlot** (matlab) Create a plot with the collected green prices with the highest value

14. **greenSpacesPlot3D** (matlab) Test the 3-D shaded surface plot for collected green prices

15. **lowestUrbanPrices** (matlab) Create a plot with the urban cells with the lowest value

16. **higherUrbanPricesPlot** (matlab) Create a plot with the urban cells with the highest value

17. **avgGreenPricesPlot** (matlab) Create a plot with the average of green prices

18. **avgUrbanPricesPlot** (matlab) Create a plot with the urban cells with the average value

19. **greenPricesPerRing** (Rings file) Plot average green price prices grouped by rings. Place Ring.txt in General folder before run the function

20. **Ring_GreenPricesPlot** (matlab) Double Plot: Plot average green price prices grouped by rings. Average green prices in EA

21. **exponentialFunction** (matlab) Return the exponential function of the non urban prices data

22. **satisfaction_protected_closeness** (matlab2) ICCS plots: 3subplot: satisfaction, number of cells protected and closeness

23. **satisfactionArea** (matlab4) Area of the satisfaction of three heuristics from the three scenarios

24. **OnOfCellsProtectedPlot** (matlab3) Plot with the number of cells protected for Online/Offline

25. **OnOfSatisfaction** (matlab3) Create a line plot with the satisfaction for Online/Offline

26. **OnOfCloseness** (matlab3) Create a plot with the closeness to CBD for Online/Offline

27. **OnOfTiming** (matlab3) Computational time for Online/Offline/Mix. **TODO.** Not really implemented

28. **OnOfLatexTableNumCells** (matlab3) Create the code in latex for a table with the satisfaction achieved by the three approaches.

29. **OnOfLatexTableSatisfaction** (matlab3) Create the code in latex for a table with the satisfaction achieved by the three approaches

30. **remainingBudget** (matlab2) Plot the function with the average remaining Budget

31. **GAOptimisationTable** *GA_Data.txt* Create a table in a figure that collects data of the EA from different scenarios. Info: Generations, Mutations, Worst, Best, Mean and SD

32. **OnOfbudget** (matlab3) *budgetONOF.txt* Plot the function with the average remaining Budget

33. **matricesCorr** *realFitness.txt* and *density.txt* Study the correlation of two matrices

34. **OnOfevolution** (matlab3) Create a line plot with the evolution for Online approach: EVOLVED NOT SELECTED, SUCCESSFUL and REJECTED.

35. **avgSalaryPlot** *Cells.txt* Create a plot with the evolution of the salary.

36. **numberCellsPlot** Create a plot with the behaviour of the cells according to their urban land-used type.

37. **urbanFailuresPlot** *InconsistencyOld2.txt* Create a boxPlot figure with the number of urban failures in function to the number of replicates gathered.

## B.3 Differences between Scenarios

### B.3.1 First attempts

- Scenario 1: Changes in the way cells are selected in GA. Fix a problem with the size of the CA in the check scenario
- Scenario 2. Add new statistic material
- Scenario 3. Fixed the growth of the city
- Scenario 4. Fixed a problem in gathering the position of people.
- Scenario 5, 6. Collected data without protection of cells. Fixed a problem in GA positions.
- Scenario 7 they depend on the distance.
  - **Fixed:** Calculating fitness values in GA. Now the protection of cells is homogeneously done.
  - **Add:** CHANGE_RATE constant = 0.2.
  - **Change:** Non-urban prices depend on the distance to CBD.
  - **Add:** Var lastUrbanised, getLastUrbanised() and setLastUrbanised()

### B.3.2 Scenario created for the ICAART journal

#### B.3.2.1 Scenario 8

- **Add:** New source of uncertainty: non-urban prices are not constant.
- **Remove:** Concept of tolerance and threshold (Genetic algorithm)
- **Add:** New form to accept a cell to be protected, urbanisation factor (Genetic algorithm)
- **Change:** Calculation of fitness (linked with urbanisation factor).
- **Fixed:** Calculation of position (Lattice, GA)

- **Fixed:** Error found in TestOptimisation::Cell::ReduceDemand

#### B.3.2.2 Scenario 9

- **Fixed:** Forest and agricultural price was swapped
- **Fixed:** *GatherData::CHANGE_RATE* was 0.8 and not 0.2.
- **Fixed:** Urban prices which gives higher values in green prices.
- **Add:** Gather statistics about cells (min, max, avg).
- **Add:** Gather statistics about the GA algorithm (max and avg fitness, mutations).
- **Fixed:** Times the mutation procedure is tried in the GA.
- **Fixed:** Statistics files were written in gather data differently than were read in GA.
- **Fixed:** Satisfaction was calculated differently in GA than in the rest of the modules.
- **Modified:** Price recent development is calculated taking all the cells in the outer annulus instead of only the last cell urbanised.

### B.3.3 Scenario created for ICCS2015

#### B.3.3.1 Scenario 10

- **Add:** More than one CBD in the simulation.
- **Add:** Capacity to choose among different fitness distances.
- **Add:** New stochastic way of generating the selections in GA.
- **Add:** Statistics gathered added automatically to *inconsistecy.txt* & *DATA_GA.txt*.
- **Add:** Information to the name of the files during all the process.
- **Add:** GA feasible & infeasible solutions.

- **Mod:** How protected cells are managed in testing
- **Fix:** Gather data shift one position the statistical data gathered
- **Add:** Checkfiles: Visualise population
- **Fix:** prices constant when they shouldn't in GA_SAT
- **Fix:** Duplicated cells to be protected
- **Fix:** One of the three cities creates cells further from CBD

#### B.3.3.2 Scenario 11

- **Modified:** Size of the lattice equal to 100.

#### B.3.3.3 Scenario 12

- **Modified:** Budget linked to population growth with the use of parameter alpha in Matlab.
  In GA_SAT: change TYPE_BUDGET = 1. Rest Lattice: TYPE_BUDGET = 1.

### B.3.4 Scenario created for UAI2015

#### B.3.4.1 Scenario 13

- **Add:** Module MultiOptimisation where GA optimisation is done online.
- **Add:** Variable Individualxxxx: times a suitable individual is searched by the GA.
- **Add:** Time total and partial gathered (Multi).
- **Fix:** Cells were not removed from the list of NON_URBAN_CELLS when they were protected. Allow a cell be protected twice.
- **Add:** Baseline.txt file to collect general information of CLO and RAN like time, CBDs...

- **Fix:** Problem found when a cell was selected for urbanised in the same turn that it was protected.

- **Add:** Initial conditions are checked in lattice.

### B.3.5 Scenario after UAI2015

#### B.3.5.1 Scenario 14

- **Fix:** Error RAN. Now more than one cell can be bought in one turn.

- **Add:** New fitness to protect high ecological values of a city

- **Mod:** Fitness calculated in each selection, not in each individual

- **Add:** Gather remaining budget to plot

- **Fix:** Error in GA when selections where populated

- **Fix:** Not allow salary < cost to select a dwelling. *avgUrbanPrices* fixed.

- **Add:** Collect real fitness to measure correlation

- **Fix:** Distance to CBD always 1 in green prices (tick=1)

- **Fix:** GA now forbids mutating an individual when it is duplicated

#### B.3.5.2 Scenario 15

Multiobjective PAES

## B.4 Future extensions

### B.4.1 Benchmarking

Perform a benchmarking comparison analysis or benchmarking between a GA approach and other kind of techniques naturally more adequate to solver sequential-decision making problems like:

- Reinforcement Learning (RL). RL was formalised by Barto (Barto et al., 1981)and it consists of a machine learning technique capable of selecting the optimal policy and representing explicitly the uncertainty.
- Simulated Annealing
- Markov Chains
- Spatial logistic regression

Goal: check if our approach is the best for this particular problem. Comparison of heuristics that depends on:

- Problem formulation.
- Parameters specifically used for the method used: determine how much time is required to complete the search (computational cost)
- Time given to search throughout the search space and performance (Pukkala and Kurttila, 2005).

### B.4.2 Improvements of the model.

- Satisfaction increases with the quality of the green area that can be measured by two factors: higher ecological value (preference for forest against agricultural land) and the extension of the protected area. Perform a pre-clustering of the areas.Measure how crowded the green areas are and penalise the satisfaction in case the area is overcrowded. Study differences in prices of buying larger extensions of land (price & satisfaction behaviour).
- Associate the green satisfaction factor and the fitness function with the individual willingness of living close to a green space.
- Adequate the salary to a normal distribution.
- Fix the problem salaries too low at the end of the simulation. (Study which part of the salary is used to pay housing expenses). Divide agents into its profile (Loibl and Toetzer, 2003).

- Implement change of residence if satisfaction is not enough. Redevelopment deactivate.
- Measure the influence of the initial scenario in the results.
- Most individuals do not have a complete knowledge of the possible set of available places. Include a stochastic factor: sometimes not the best areas are developed.
- Search better behaviour for migration.
- A non-homogeneous distribution of resources.
- Array that initialise CBD with starting time.
- Think in an intelligent heuristic like CLO for protection of green areas

## B.5 Notes

Directions in all modules:

$$dir = (x * size) + y$$

In CLO searchGreenSpaces is triggered by municipality and not by updateAggregate.

Urban cells in class *city* are considered static, however each cell knows in *CITY_ID*, the city it belongs.

# References


Acharya, Ballabh Prasad. Policy of land acquisition and development: Analysis of an Indian experience. *Third World Planning Review*, 9(2):99, 1987.

Acharya, Gayatri and Bennett, Lynne Lewis. Valuing open space and land-use patterns in urban watersheds. *The Journal of Real Estate Finance and Economics*, 22(2-3):221–237, 2001.

Aerts, Jeroen CJH and Heuvelink, Gerard BM. Using simulated annealing for resource allocation. *International Journal of Geographical Information Science*, 16 (6):571–587, 2002.

Ageev, Alexander A and Sviridenko, Maxim I. Approximation algorithms for maximum coverage and max cut with given sizes of parts. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 17–30. Springer, 1999.

Agogino, Adrian K and Tumer, Kagan. A multiagent approach to managing air traffic flow. *Autonomous Agents and Multi-Agent Systems*, 24(1):1–25, 2012.

Aguilera, Francisco; Valenzuela, Luis M, and Botequilha-Leitão, André. Landscape metrics in the analysis of urban land use patterns: A case study in a Spanish metropolitan area. *Landscape and Urban Planning*, 99(3):226–238, 2011.

Ahern, Jack. Ecopolis; strategies for ecologically sound urban development. *Landscape and Urban Planning*, 3(36):232–233, 1995.

Aizawa, Akiko N and Wah, Benjamin W. Scheduling of genetic algorithms in a noisy environment. *Evolutionary Computation*, 2(2):97–122, 1994.

Akın, Anıl; Clarke, Keith C, and Berberoglu, Suha. The impact of historical exclusion on the calibration of the SLEUTH urban growth model. *International Journal of Applied Earth Observation and Geoinformation*, 27:156–168, 2014.

Al-Kheder, Sharaf; Wang, Jing, and Shan, Jie. Fuzzy inference guided cellular automata urban-growth modelling using multi-temporal satellite images. *International Journal of Geographical Information Science*, 22(11-12):1271–1293, 2008.

Albareda-Sambola, Maria; Fernández, Elena; Hinojosa, Yolanda, and Puerto, Justo. The multi-period incremental service facility location problem. *Computers & Operations Research*, 36(5):1356–1375, 2009.

Alberti, Marina and Marzluff, John M. Ecological resilience in urban ecosystems: linking urban patterns to human and ecological functions. *Urban Ecosystems*, 7 (3):241–265, 2004.

Allen, Peter M and Sanglier, Michele. A dynamic model of growth in a central place system. *Geographical Analysis*, 11(3):256–272, 1979.

Allison, Lincoln. *Environmental Planning: A Political and Philosophical Analysis*. Allen and Unwin (London), 1975.

Alonso, William. *Location and Land Use. Toward a General Theory of Land Rent*. Harvard University Press, Cambridge, USA, January 1964. doi:10.4159/harvard.9780674730854. URL `http://dx.doi.org/10.4159/harvard.9780674730854`.

Alterman, Rachelle. Between Privatization and Continued National Ownership: A Future Land Policy for Israel. *The Floersheimer Institute for Policy Studies, Jerusalem (Hebrew)*, 1999.

Amato, Christopher; Chowdhary, Girish; Geramifard, Alborz; Ure, N Kemal, and Kochenderfer, Mykel J. Decentralized control of partially observable Markov decision processes. In *IEEE 52nd Annual Conference on Decision and Control (CDC)*, pages 2398–2405. IEEE, 2013.

Anagnostopoulos, KP and Mamanis, Georgios. A portfolio optimization model with three objectives and discrete variables. *Computers & Operations Research*, 37(7): 1285–1297, 2010.

Andersson, Claes; Rasmussen, Steen, and White, Roger. Urban settlement transitions. *Environment and Planning B*, 29(6):841–866, 2002.

Antoni, Jean-Philippe. Urban sprawl modelling: a methodological approach. *Cybergeo: European Journal of Geography*, 2002.

Arnberger, Arne. Recreation use of urban forests: An inter-area comparison. *Urban Forestry & Urban Greening*, 4(3):135–144, 2006.

Arnberger, Arne. Urban densification and recreational quality of public urban green spaces—a Viennese case study. *Sustainability*, 4(4):703–720, 2012.

Arnberger, Arne and Brandenburg, Christiane. Past on-site experience, crowding perceptions, and use displacement of visitor groups to a peri-urban national park. *Environmental Management*, 40(1):34–45, 2007.

Arnold, Dirk V. *Noisy Optimization with Evolution Strategies*, volume 8. Springer Science+Business Media, 2002.

Arnold, Dirk V and Beyer, Hans-Georg. Local performance of the ($\mu/\mu$I, $\lambda$)-ES in a noisy environment. *Foundations of Genetic Algorithms*, 6:127–141, 2001.

Arnold, Dirk V and Beyer, Hans-Georg. A comparison of evolution strategies with other direct search methods in the presence of noise. *Computational Optimization and Applications*, 24(1):135–159, 2003.

Arnott, Richard J and Lewis, Frank D. The Transition of Land to Urban Use. *The Journal of Political Economy*, 87(1):161–169, 1979.

Arsanjani, Jamal Jokar; Helbich, Marco; Kainz, Wolfgang, and Boloorani, Ali Darvishi. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, 21:265–275, 2013.

Artel, Arsun; Teymour, Fouad; North, Michael, and Cinar, Ali. A multi-agent approach using perceptron-based learning for robust operation of distributed chemical reactor networks. *Engineering Applications of Artificial Intelligence*, 24 (6):1035–1045, 2011.

Axelrod, Robert M. *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton University Press, 1997.

Babbar, Meghna; Lakshmikantha, Ashvin, and Goldberg, David E. A modified NSGA-II to solve noisy multiobjective problems. In *Genetic and Evolutionary Computation Conference. Late-Breaking Papers (GECCO 2003)*, volume 2723, pages 21–27, 2003.

Bäck, Thomas; Hoffmeister, Frank, and Schwefel, Hans-Paul. A survey of evolution strategies. In *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 2–9, 1991.

Bäck, Thomas; Fogel, David B., and Michalewicz, Zbigniew. *Handbook of Evolutionary Computation*. Bristol: Institute of Physics, 1997. URL `http://148.201.96.14/dc/ver.aspx?ns=000155042`.

Badri, Masood A; Mortagy, Amr K, and Alsayed, Colonel Ali. A multi-objective model for locating fire stations. *European Journal of Operational Research*, 110(2): 243–260, 1998.

Baker, James E. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms*, pages 14–21, 1987.

Balling, Richard J; Taber, John T; Brown, Michael R, and Day, Kirsten. Multiobjective Urban Planning using Genetic Algorithm. *Journal of Urban Planning and Development*, 125(2):86–99, June 1999. URL `http://worldcat.org/issn/07339488`.

Ballou, Ronald H. Dynamic warehouse location analysis. *Journal of Marketing Research*, pages 271–276, 1968.

Balmer, Michael; Cetin, Nurhan; Nagel, Kai, and Raney, Bryan. Towards truly agent-based traffic and mobility simulations. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 60–67. IEEE Computer Society, 2004.

Balvanera, Patricia; Kremen, Claire, and Martínez-Ramos, Miguel. Applying community structure analysis to ecosystem function: examples from pollination and carbon storage. *Ecological Applications*, 15(1):360–375, 2005.

Barker, Terry. *Disaggregation in Econometric Modelling (Routledge Revivals)*. Routledge, 2014.

Barredo, José I; Kasanko, Marjo; McCormick, Niall, and Lavalle, Carlo. Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64(3):145–160, 2003.

Barredo, José I; Demicheli, Luca; Lavalle, Carlo; Kasanko, Marjo, and McCormick, Niall. Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria. *Environment and Planning B*, 31(1):65–84, 2004.

Barreira-González, Pablo; Gómez-Delgado, Montserrat, and Aguilera-Benavente, Francisco. From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, 54:119–131, 2015.

Barthelemy, J-FM and Haftka, Raphael T. Approximation concepts for optimum structural design—a review. *Structural and Multidisciplinary Optimization*, 5(3): 129–144, 1993.

Barto, Andrew G.; Sutton, Richard S., and Brouwer, Peter S. Associative search network: A reinforcement learning associative memory. *Biological Cybernetics*, 40: 201–211, 1981. ISSN 0340-1200. URL `http://dx.doi.org/10.1007/BF00453370`. 10.1007/BF00453370.

Barton, Jo and Pretty, Jules. What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. *Environmental Science & Technology (ACS Publications)*, 44(10):3947–3955, 2010.

Barton, Russell R. Metamodeling: a state of the art review. In *Proceedings of the 1994 Winter Simulation Conference, Lake Buena Vista, Florida, USA*, pages 237–244. IEEE, December 1994.

Batty, Michael. Progress, success, and failure in urban modelling. *Environment and Planning A*, 11(8):863–878, 1979.

Batty, Michael. *Cities and Complexity: Understanding Cities Through Cellular Automata, Agent-Based Models and Fractals.* The MIT Press, 2007.

Batty, Michael. Urban modeling. *International Encyclopedia of Human Geography. Oxford, UK: Elsevier*, 2009.

Batty, Michael. *The New Science of Cities.* The MIT Press, 2013.

Batty, Michael and Longley, Paul A. *Fractal Cities: a Geometry of Form and Function.* Academic Press, 1994.

Batty, Michael and Torrens, Paul M. Modelling and prediction in a complex world. *Futures*, 37(7):745 – 766, 2005. ISSN 0016-3287. doi:DOI: 10.1016/j.futures.2004.11.003. URL `http://www.sciencedirect.com/science/article/B6V65-4FRKVKW-1/2/70a318cd31a6f252b0dd1b15534d0c71`. Complexity and the limits of knowledge.

Batty, Michael and Xie, Yichun. From cells to cities. *Environment and Planning B: Planning and Design*, 21(7):S31–S48, 1994.

Batty, Michael; Couclelis, Helen, and Eichen, Marc. Urban systems as cellular automata. *Environment and Planning B: Planning and Design*, 24(2):159–164, 1997.

Batty, Michael; Xie, Yichun, and Sun, Zhanli. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems*, 23(3): 205–233, 1999.

Bedimo-Rung, Ariane L; Mowen, Andrew J, and Cohen, Deborah A. The significance of parks to physical activity and public health: a conceptual model. *American Journal of Preventive Medicine*, 28(2):159–168, 2005.

Bellman, Richard. *Dynamic Programming*. Princeton University Press, Princeton, NJ., 1957.

Belyea, Curtis M and Terando, Adam J. Designing Sustainable Landscapes. *Biodiversity and Spatial Information Center, NC State University, Raleigh*, 2010.

Benenson, Itzhak and Torrens, Paul M. *Geosimulation: Automata-based Modeling of Urban Phenomena*. John Wiley & Sons, 2004.

Benenson, Itzhak and Torrens, Paul M. System Theory, Geography, and Urban Modeling. *Geosimulation: Automata-Based Modeling of Urban Phenomena*, pages 47–90, 2006.

Bengochea Morancho, Aurelia. A hedonic valuation of urban green areas. *Landscape and Urban Planning*, 66(1):35–41, 2003.

Bennett, David A and Tang, W. Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. *International Journal of Geographical Information Science*, 20(9):1039–1066, 2006.

Bernatzky, Aloys. The effects of trees on the urban climate. *Trees in the 21st Century. Academic Publishers, Berkhamster*, pages 59–76, 1983.

Berry, Brian JL. *The Changing Shape of Metropolitan America*. Ballinger Publishing Company, 1977.

Berry, Brian JL and Kim, Hak-Min. Challenges to the monocentric model. *Geographical Analysis*, 25(1):1–4, 1993.

Bert, Federico E; Podestá, Guillermo P; Rovere, Santiago L; Menéndez, Ángel N; North, Michael; Tatara, Eric; Laciana, Carlos E; Weber, Elke, and Toranzo, Fernando Ruiz. An agent based model to simulate structural and land use changes in agricultural systems of the argentine pampas. *Ecological Modelling*, 222(19): 3486–3499, 2011.

Bertaud, Alain. The spatial organization of cities: Deliberate outcome or unforeseen consequence? *Working Papers, Institute of Urban and Regional Development (IURD), University of California Berkeley*, 1, 2004.

Bettinger, Pete; Sessions, John; Chung, Woodam; Graetz, David, and Boston, Kevin. Eight Heuristic Planning Techniques Applied to Three Increasingly Difficult Wildlife Planning Problems: A Summary. In Arthaud, Greg J. and Barrett, Tara M., editors, *Systems Analysis in Forest Resources: Proceedings of the Eighth Symposium, Snowmass Village, Colorado, U.S.A.*, pages 249–257. Springer Netherlands, Dordrecht, September 2003. ISBN 978-94-017-0307-9. doi:10.1007/978-94-017-0307-9_24. URL `http://dx.doi.org/10.1007/978-94-017-0307-9_24`.

Beyer, Hans-Georg. Toward a theory of evolution strategies: On the benefits of sex—the ($\mu/\mu$, $\lambda$) theory. *Evolutionary Computation*, 3(1):81–111, 1995.

Beyer, Hans-Georg. Evolutionary algorithms in noisy environments: theoretical issues and guidelines for practice. *Computer Methods in Applied Mechanics and Engineering*, 186(2):239–267, 2000.

Beyer, Hans-Georg; Olhofer, Markus, and Sendhoff, Bernhard. On the Behavior of ($\mu/\mu1$, $\lambda$)-ES Optimizing Functions Disturbed by Generalized Noise. In *Foundations of Genetic Algorithms (FOGA), Torremolinos, Spain*, pages 307–328, September 2002.

Biles, John A. GenJam: A genetic algorithm for generating jazz solos. In *Proceedings of the International Computer Music Conference (ICMC)*, volume 94, pages 131–137. International Computer Music Association, 1994.

Bixler, Robert D and Floyd, Myron F. Nature is scary, disgusting, and uncomfortable. *Environment and Behavior*, 29(4):443–467, 1997.

Björk, Jonas; Albin, Maria; Grahn, Patrik; Jacobsson, Hanna; Ardö, Jonas; Wadbro, John; Östergren, Per-Olof, and Skärbäck, Erik. Recreational values of the natural environment in relation to neighbourhood satisfaction, physical activity, obesity and wellbeing. *Journal of Epidemiology and Community Health*, 62(4):62–e2, 2008.

Bodin, Maria and Hartig, Terry. Does the outdoor environment matter for psychological restoration gained through running? *Psychology of Sport and Exercise*, 4 (2):141–153, 2003.

Boero, Riccardo and Squazzoni, Flaminio. Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *Journal of Artificial Societies and Social Simulation*, 8(4), 2005.

Bogle, Gib and Dunbar, P Rod. Agent-based simulation of T-cell activation and proliferation within a lymph node. *Immunology & Cell Biology*, 88(2):172–179, 2010.

Bolitzer, Benjamin and Netusil, Noelwah R. The impact of open spaces on property values in Portland, Oregon. *Journal of Environmental Management*, 59(3):185–193, 2000.

Bolund, Per and Hunhammar, Sven. Ecosystem services in urban areas. *Ecological Economics*, 29(2):293–301, 1999.

Bone, Christopher and Dragicevic, Suzana. Incorporating spatio-temporal knowledge in an Intelligent Agent Model for natural resource management. *Landscape and Urban Planning*, 96(2):123 – 133, 2010. ISSN 0169-2046. doi:DOI: 10.1016/j.landurbplan.2010.03.002. URL `http://www.sciencedirect.com/science/article/pii/S0169204610000484`.

Bonini, Charles P. *Simulation of Information and Decision System in the Firm*. Prentice-Hall, 1963.

Boone, Christopher G; Buckley, Geoffrey L; Grove, J Morgan, and Sister, Chona. Parks and people: An environmental justice inquiry in Baltimore, Maryland. *Annals of the Association of American Geographers*, 99(4):767–787, 2009.

Borgström, Sara T; Elmqvist, Thomas; Angelstam, Per, and Alfsen-Norodom, Christine. Scale mismatches in management of urban landscapes. *Ecology and Society*, 11(2):1–16, 2006.

Bovy, Piet and Jansen, GRM. Network aggregation effects upon equilibrium assignment outcomes: An empirical investigation. *Transportation Science*, 17(3): 240–262, 1983.

Bow, C Jennifer D; Waters, Nigel M; Faris, Peter D; Seidel, Judy E; Galbraith, P Diane; Knudtson, Merril L, and Ghali, William A. Accuracy of city postal code coordinates as a proxy for location of residence. *International Journal of Health Geographics*, 3(1):5, 2004.

Bowler, Diana E; Buyung-Ali, Lisette; Knight, Teri M, and Pullin, Andrew S. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, 97(3):147–155, 2010a.

Bowler, Diana E; Buyung-Ali, Lisette M; Knight, Teri M, and Pullin, Andrew S. A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, 10(1):456, 2010b.

Boyd, James. Location, location, location: the geography of ecosystem services. *Resources for the Future*, pages 11–15, 2008.

Brander, Luke M and Koetse, Mark J. The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. *Journal of Environmental Management*, 92(10):2763–2773, 2011.

Branke, Jürgen. Creating robust solutions by means of evolutionary algorithms. In *Parallel Problem Solving from Nature—PPSN V*, pages 119–128. Springer, 1998.

Branke, Jürgen and Schmidt, Christian. Selection in the presence of noise. In *Proceedings of the Genetic and Evolutionary Computation (GECCO), Chicago, IL, USA*, pages 766–777. Springer, 2003.

Branke, Jürgen; Schmidt, Christian, and Schmec, Hartmut. Efficient fitness estimation in noisy environments. In *Proceedings of Genetic and Evolutionary Computation (GECCO), San Francisco, California, USA*. Citeseer, 2001.

Brauman, Kate A; Daily, Gretchen C; Duarte, T Ka'eo, and Mooney, Harold A. The nature and value of ecosystem services: an overview highlighting hydrologic services. *Annual Review of Environment and Resources*, 32:67–98, 2007.

Breffle, William S; Morey, Edward R, and Lodder, Tymon S. Using contingent valuation to estimate a neighbourhood's willingness to pay to preserve undeveloped urban land. *Urban Studies*, 35(4):715–727, 1998.

Brimberg, Jack and Mladenovic, Nenad. Solving the continuous location-allocation problem with Tabu search. *Studies in Locational Analysis*, 8(41):23–32, 1996.

Broomhall, Melissa Helen. Study of the availability and environmental quality of urban open space used for physical activity. *Master of Public Health Dissertation. Department of Public Health. University of Western Australia, Perth*, 1996.

Brown, Daniel G. and Robinson, Derek T. Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl. *Ecology and Society*, 11 (1), 2006. URL `http://www.ecologyandsociety.org/vol11/iss1/art46/`.

Brown, Daniel G; Page, Scott E; Riolo, Rick, and Rand, William. Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling & Software*, 19(12):1097–1109, 2004.

Brown, Daniel G; Page, Scott; Riolo, Rick; Zellner, Moira, and Rand, William. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19(2):153–174, 2005.

Brown, Daniel G.; Aspinall, Richard, and Bennett, David A. Landscape Models and Explanation in Landscape Ecology. A Space for Generative Landscape Science? *The Professional Geographer*, 58:369–382(14), November 2006. doi:10.1111/j.1467-9272.2006.00575.x.

Brueckner, Jan K; Thisse, Jacques-Francois, and Zenou, Yves. Why is central Paris rich and downtown Detroit poor?: An amenity-based theory. *European Economic Review*, 43(1):91–107, 1999.

Brunt, Paul and Courtney, Paul. Host perceptions of sociocultural impacts. *Annals of Tourism Research*, 26(3):493–515, 1999.

Büche, Dirk; Stoll, Peter, and Koumoutsakos, Petros. An evolutionary algorithm for multi-objective optimization of combustion processes. *Center for Turbulence Research Annual Research Briefs*, 2001:231–239, 2001.

Büche, Dirk; Stoll, Peter; Dornberger, Rolf, and Koumoutsakos, Petros. Multiobjective evolutionary algorithm for the optimization of noisy combustion processes. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 32(4):460–473, 2002.

Bui, Lam T; Abbass, Hussein A, and Essam, Daryl. Fitness inheritance for noisy evolutionary multi-objective optimization. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation (GECCO), Washington, DC, USA*, pages 779–785. ACM, June 2005.

Burgess, Jacquelin; Harrison, Carolyn M, and Limb, Melanie. People, parks and the urban green: a study of popular meanings and values for open spaces in the city. *Urban Studies*, 25(6):455–473, 1988.

Burke, Jessica; O'Campo, Patricia; Salmon, Christina, and Walker, Renee. Pathways connecting neighborhood influences and mental well-being: socioeconomic position and gender differences. *Social Science & Medicine*, 68(7):1294–1304, 2009.

Byomkesh, Talukder; Nakagoshi, Nobukazu, and Dewan, Ashraf. Urbanization and green space dynamics in Greater Dhaka, Bangladesh. *Landscape and Ecological Engineering*, pages 1–14, 2010. ISSN 1860-1871. URL `http://dx.doi.org/10.1007/s11355-010-0147-7`. 10.1007/s11355-010-0147-7.

Cantú-Paz, Erick. Adaptive sampling for noisy problems. In *Proceedings of the*

*Genetic and Evolutionary Computation (GECCO), Seattle, WA, USA*, pages 947–958. Springer, 2004.

Cao, Kai; Batty, Michael; Huang, Bo; Liu, Yan; Yu, Le, and Chen, Jiongfeng. Spatial multi-objective land use optimization: extensions to the non-dominated sorting genetic algorithm-II. *International Journal of Geographical Information Science*, 25(12):1949–1969, 2011.

Cao, Kai; Huang, Bo; Li, Manchun, and Li, Wenwen. Calibrating a cellular automata model for understanding rural–urban land conversion: A Pareto front-based multi-objective optimization approach. *International Journal of Geographical Information Science*, 28(5):1028–1046, 2014.

Capozza, Dennis R and Helsley, Robert W. The fundamentals of land prices and urban growth. *Journal of Urban Economics*, 26(3):295–306, 1989.

Carr, Stephen. *Public Space*. Cambridge University Press, 1992.

Caruso, Geoffrey; Peeters, Dominique; Cavailhès, Jean, and Rounsevell, Mark. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*, 37(5):542–567, 2007.

Cavailhès, Jean and Wavresky, Pierre. Urban influences on periurban farmland prices. *European Review of Agricultural Economics*, 30(3):333–357, 2003.

Ceballos Lascurain, Hector. *Tourism, Ecotourism, and Protected Areas: The State of Nature-based Tourism Around the World and Guidelines for its Development*. Gland, Switzerland, Cambridge, UK: IUCN and The World Conservation Union, 1996.

Cecchini, Arnaldo and Rizzi, Paola. The reasons why cellular automata are a useful tool in the working-kit for the new millennium urban planner in governing the territory. *Proceedings of the 7th International Conference on Computers in Urban Planning and Urban Management (CUPUM). University of Hawaii, Honolulu.*, 2001.

Chambers, M and Mount-Campbell, CA. Process optimization via neural network metamodeling. *International Journal of Production Economics*, 79(2):93–100, 2002.

Chardaire, Pierre; Sutter, Alain, and Costa, Marie-Christine. Solving the dynamic facility location problem. *Networks*, 28(2):117–124, 1996.

Chen, Hudong; Chen, Shiyi; Doolen, Gary, and Lee, YC. Simple lattice gas models for waves. *Complex Systems*, 2:259–267, 1988.

Chen, Jln; Gong, Peng; He, Chunyang; Luo, Wei; Tamura, Masayukl, and Shi, Peijun. Assessment of the urban development plan of Beijing by using a CA-based urban growth model. *Photogrammetric Engineering & Remote Sensing*, 68(10):1063–1072, 2002.

Chen, Wendy Y and Jim, CY. Cost–benefit analysis of the leisure value of urban greening in the new Chinese city of Zhuhai. *Cities*, 25(5):298–309, 2008.

Cheng, Jianquan. *Modelling Spatial and Temporal Urban Growth*. Utrecht University, 2003.

Cheshire, Paul and Sheppard, Stephen. Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics*, 60(3):357–382, 1998.

Chicoine, David L. Farmland values at the urban fringe: an analysis of sale prices. *Land Economics*, pages 353–362, 1981.

Chiesura, Anna. The role of urban parks for the sustainable city. *Landscape and Urban Planning*, 68(1):129–138, 2004.

Chingcuanco, Franco and Miller, Eric J. A microsimulation model of urban energy use: Modelling residential space heating demand in ILUTE. *Computers, Environment and Urban Systems*, 36(2):186–194, 2012.

Church, Richard L. Geographical information systems and location science. *Computers & Operations Research*, 29(6):541–562, 2002.

Ciaian, Pavel; Kancs, d'Artis; Swinnen, Jo; Van Herck, Kristine, and Vranken, Liesbet. Sales Market Regulations for Agricultural Land in EU Member States and Candidate Countries. Technical report, Centre for European Policy Studies, 2012.

Ciesielski, Vic and Scerri, Paul. Real time genetic scheduling of aircraft landing times. In *The 1998 IEEE International Conference on Evolutionary Computation(ICEC98),NY,USA*, pages 360–364. IEEE, 1998.

Clark, Colin. Urban Population Densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):pp. 490–496, 1951. ISSN 00359238. URL `http://www.jstor.org/stable/2981088`.

Clarke, K and Hoppen, S. A self-modifying cellular automaton model of historical urbaniztation in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, 24(2):247–261, 1997.

Clarke, Keith C. A decade of cellular urban modeling with SLEUTH: Unresolved issues and problems. *Planning Support Systems for Cities and Regions*, 3:47–60, 2008.

Clarke, Keith C and Gaydos, Leonard J. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12(7): 699–714, 1998.

Clarke, Keith C; Brass, James A, and Riggan, Philip J. A cellular automation model of wildfire propagation and extinction. *Photogrammetric Engineering & Remote Sensing*, 60(11):1355–1367, 1994.

Clarke-Lauer, MD and Clarke, Keith C. Evolving Simulation Modeling: Calibrating SLEUTH Using a Genetic Algorithm. In *Proceedings of the 11th International Conference on Geo Computation, University College London 20th–22nd July*, 2011.

Coello, Carlos A Coello. An updated survey of evolutionary multiobjective optimization techniques: State of the art and future trends. In *Proceedings of the Congress on Evolutionary Computation*, volume 1, pages 3–13, 1999.

Coello, Carlos A Coello. An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys (CSUR)*, 32(2):109–143, 2000.

Coello, Carlos A Coello and Pulido, Gregorio Toscano. A micro-genetic algorithm for multiobjective optimization. In *Evolutionary Multi-criterion Optimization*, pages 126–140. Springer, 2001.

Coley, Rebekah Levine; Sullivan, William C, and Kuo, Frances E. Where does community grow? The social context created by nature in urban public housing. *Environment and Behavior*, 29(4):468–494, 1997.

Colwell, Peter F and Munneke, Henry J. The structure of urban land prices. *Journal of Urban Economics*, 41(3):321–336, 1997.

Colwell, Peter F and Munneke, Henry J. Land prices and land assembly in the CBD. *The Journal of Real Estate Finance and Economics*, 18(2):163–180, 1999.

Colwell, Peter F and Sirmans, Clemon F. Area, time, centrality and the value of urban land. *Land Economics*, pages 514–519, 1978.

Conway, John. Game of Life. *Scientific American*, 223:120 –123, 1970.

Cooke, Philip. *Theories of Planning and Spatial Development.* Hutchinson London, 1983.

Cooper, Leon. Location-allocation problems. *Operations Research*, 11(3):331–343, 1963.

Cornuejols, Gerard; Fisher, Marshall L, and Nemhauser, George L. Exceptional paper-location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science*, 23(8):789–810, 1977.

Cornuejols, Gerard; Nemhauser, George L, and Wolsey, Laurence A. Worst-case and probabilistic analysis of algorithms for a location problem. *Operations Research*, 28(4):847–858, 1980.

Costanza, Robert; d'Arge, Ralph; De Groot, Rudolf; Farber, Stephen; Grasso, Monica; Hannon, Bruce; Limburg, Karin; Naeem, Shahid; O'neill, Robert V; Paruelo, Jose; Raskin, R.G.; Sutton, P., and van den Belt, M. The value of the world's ecosystem services and natural capital. *Ecological Economics*, 25(1):3–15, 1998. doi:10.1016/S0921-8009(98)00020-2.

Couclelis, Helen. Cellular worlds: a framework for modeling micro - macro dynamics. *Environment and Planning A*, 15(5):585 – 596, 19, June 1984. doi:10.1068/a170585. URL `http://envplan.com/abstract.cgi?id=a170585`.

Couclelis, Helen. From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B: Planning and Design*, 24(2):165–174, 1997. URL `http://econpapers.repec.org/RePEc:pio:envirb:v:24:y:1997:i:2:p:165-174`.

Crane, Philip R; Wyeth, Stephanie J; Brough, Mark K, and Spencer, Anna. Children in inner-city suburbia: The case of New Farm, Brisbane. In *In Proceedings Creating Child-friendly Cities Conference, Sydney*. The Australian Research Alliance for Children and Youth (ARACY), October 2006.

Creutz, Michael. Deterministic ising dynamics. *Annals of Physics*, 167(1):62–72, 1986.

Crociani, Luca; Piazzoni, Andrea, and Vizzari, Giuseppe. Adaptive Hybrid Agents for Tactical Decisions in Pedestrian Environments. In *Proceedings of WOA - XVI Workshop 'From Objects to Agents', Naples, Italy*, June 2015.

Crowe, Kevin and Nelson, John. An indirect search algorithm for harvest-scheduling under adjacency constraints. *Forest Science*, 49(1):1–11, 2003.

Crucifix, Michel and Rougier, Jonathan. On the use of simple dynamical systems for climate predictions. *The European Physical Journal-Special Topics*, 174(1):11–31, 2009.

da Gama, F Saldanha. *Models and algorithms for the dynamic facility location problem*. PhD thesis, Faculty of Sciences, University of Lisbon, Portugal.(In Portuguese), 2002.

Dadvand, Payam; de Nazelle, Audrey; Figueras, Francesc; Basagaña, Xavier; Su, Jason; Amoly, Elmira; Jerrett, Michael; Vrijheid, Martine; Sunyer, Jordi, and Nieuwenhuijsen, Mark J. Green space, health inequality and pregnancy. *Environment International*, 40:110–115, 2012.

Dadvand, Payam; Wright, John; Martinez, David; Basagaña, Xavier; McEachan, Rosemary RC; Cirach, Marta; Gidlow, Christopher J; de Hoogh, Kees; Gražulevičienė, Regina, and Nieuwenhuijsen, Mark J. Inequality, green spaces, and pregnant women: Roles of ethnicity and individual and neighbourhood socioeconomic status. *Environment International*, 71:101–108, 2014.

Daganzo, Carlos F. *Logistics Systems Analysis*. Springer Science+Business Media, 2005.

Daneke, Christian. *Modellierung von Stadtentwicklungsprozessen am Fallbeispiel Hamburg: unter Berücksichtigung stadtklimatologischer Aspekte*. Franz Steiner Verlag, 2013.

Darwin, Charles. *On the Origin of Species by Means of Natural Selections: Or the Preservation of Favoured Races in the Struggle for Life*. Murray, 1861.

Daskin, Mark S. Network and Discrete Location: Models, Algorithms and Applications. *Journal of the Operational Research Society*, 48:763–763, July 1997. URL http://www.ingentaconnect.com/content/pal/01605682/1997/00000048/00000007/2600828.

Daskin, Mark S. What you should know about location modeling. *Naval Research Logistics (NRL)*, 55(4):283–294, 2008.

Daskin, Mark S. *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley & Sons, 2011.

Daskin, Mark S; Hopp, Wallace J, and Medina, Benjamin. Forecast horizons and dynamic facility location planning. *Annals of Operations Research*, 40(1):125–151, 1992.

Datta, Dilip; Deb, Kalyanmoy, and Fonseca, Carlos M. Multi-Objective Evolutionary Algorithm for University Class Timetabling Problem. In Dahal, Keshav P.; Tan, Kay Chen, and Cowling, Peter I., editors, *Evolutionary Scheduling*, pages 197–236. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-48584-1. doi:10.1007/978-3-540-48584-1_8.

Datta, Dilip; Fonseca, Carlos M, and Deb, Kalyanmoy. A multi-objective evolutionary algorithm to exploit the similarities of resource allocation problems. *Journal of Scheduling*, 11(6):405–419, 2008.

Davis, Lawrence. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold Company, February 1991.

Deb, Kalyanmoy. *Multi-Objective Optimization Using Evolutionary Algorithms*, volume 16. John Wiley & Sons, 2001.

Deb, Kalyanmoy; Agrawal, Samir; Pratap, Amrit, and Meyarivan, Tanaka. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Lecture Notes in Computer Science*, 1917:849–858, 2000.

Deitz, Richard. A joint model of residential and employment location in urban areas. *Journal of Urban Economics*, 44(2):197–215, 1998.

Dekle, Jamie; Lavieri, Mariel S; Martin, Erica; Emir-Farinas, Hülya, and Francis, Richard L. A Florida county locates disaster recovery centers. *Interfaces*, 35(2): 133–139, 2005.

del Saz Salazar, Salvador. *Valoración Social de la Zona de Ocio del Moll de Costa (Puerto de Castellón)*. Civitas, S.A., 2000.

del Saz Salazar, Salvador and García, Menendéz Leandro. Estimating the non-market benefits of an urban park: Does proximity matter? *Land Use Policy*, 24(1): 296–305, 2007.

del Saz Salazar, Salvador; García Menéndez, Leandro, and Palau Navarro, José Manuel. *Los Beneficios Sociales de la Remodelación Urbanística de la Fachada Litoral de Valencia: un Estudio de Valoración Contingente*. Civitas, S.A., 1999.

Dendrinos, Dimitrios S and Mullally, Henry. *Urban Evolution: Studies in the Mathematical Ecology of Cities*. Oxford University Press Oxford, 1985.

Devisch, Oswald TJ; Timmermans, Harry JP; Arentze, Theo A, and Borger, Aloys WJ. An agent-based model of residential choice dynamics in nonstationary housing markets. *Environment and Planning A*, 41(8):1997 – 2013, 2009.

Di Pietro, Anthony; While, Lyndon, and Barone, Luigi. Applying evolutionary algorithms to problems with noisy, time-consuming fitness functions. In *Congress on Evolutionary Computation, CEC2004*, volume 2, pages 1254–1261. IEEE, 2004.

Ditton, Robert B and Sutton, Stephen G. Substitutability in recreational fishing. *Human Dimensions of Wildlife*, 9(2):87–102, 2004.

Doick, Kieron J; Peace, Andrew, and Hutchings, Tony R. The role of one large greenspace in mitigating London's nocturnal urban heat island. *Science of The Total Environment*, 493:662–671, 2014.

Donovan, Geoffrey H and Butry, David T. The value of shade: estimating the effect of urban trees on summertime electricity use. *Energy and Buildings*, 41(6):662–668, 2009.

Ducheyne, Els I; De Wulf, Robert R, and De Baets, Bernard. A spatial approach to forest-management optimization: linking GIS and multiple objective genetic algorithms. *International Journal of Geographical Information Science.*, 20:917–928, 2006. URL `http://www.informaworld.com/10.1080/13658810600711287`.

Durmaz, Engin; Aras, Necati, and Altınel, İ Kuban. Discrete approximation heuristics for the capacitated continuous location–allocation problem with probabilistic customer locations. *Computers & Operations Research*, 36(7):2139–2148, 2009.

Ehlert, Patrick AM and Rothkrantz, Leon JM. Microscopic traffic simulation with reactive driving agents. In *Proceedings of IEEE Intelligent Transportation Systems*, pages 860–865, 2001.

Eisler, Anna D; Eisler, Hannes, and Yoshida, Mitsuo. Perception of human ecology: cross-cultural and gender comparisons. *Journal of Environmental Psychology*, 23 (1):89–101, 2003.

Engelen, Guy; White, Roger, and Nijs, Ton. Environment Explorer: spatial support system for the integrated assessment of socio-economic and environmental policies in the Netherlands. *Integrated Assessment*, 4(2):97–105, 2003.

Epstein, Joshua M and Axtell, Robert. *Growing Artificial Societies: Social Science from the Bottom up*. Brookings Institution Press, 1996.

Epstein, Leonard H; Paluch, Rocco A; Gordy, Constance C, and Dorn, Joan. Decreasing sedentary behaviors in treating pediatric obesity. *Archives of Pediatrics and Adolescent Medicine*, 154(3):220–226, 2000.

Erlenkotter, Donald. A comparative study of approaches to dynamic location problems. *European Journal of Operational Research*, 6(2):133–143, 1981.

Eshelman, Larry J and Schaffer, J David. Preventing Premature Convergence in Genetic Algorithms by Preventing Incest. In *Proceedings of the Fourth International Conference on Genetic Algorithms, University of California, San Diego, USA*, volume 91, pages 115–122, July 1991.

Ettema, Dick. A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems*, 35(1):1–11, 2011.

Faghri, Ardeshir; Lang, Adam; Hamad, Khaled, and Henck, Heather. Integrated knowledge-based geographic information system for determining optimal location of park-and-ride facilities. *Journal of Urban Planning and Development*, 128(1): 18–41, 2002.

Farahani, Reza Zanjirani; Hekmatfar, Masoud; Fahimnia, Behnam, and Kazemzadeh, Narges. Hierarchical facility location problem: Models, classifications, techniques, and applications. *Computers & Industrial Engineering*, 68:104–117, 2014.

Farhan, Bilal and Murray, Alan T. Siting park-and-ride facilities using a multi-objective spatial optimization model. *Computers & Operations Research*, 35(2): 445–456, 2008.

Feige, Uriel. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

Ferber, Jacques. *Multi-agent Systems: an Introduction to Distributed Artificial Intelligence*, volume 1. Addison-Wesley Reading, 1999.

Ferreira, Giordano BS.; Vargas, Patricia A., and Oliveira, Gina MB. An Improved Cellular Automata-Based Model for Robot Path-Planning. In Mistry, Michael; Leonardis, Aleš; Witkowski, Mark, and Melhuish, Chris, editors, *Proceedings of Advances in Autonomous Robotics Systems: 15th Annual Conference TAROS, Birmingham, UK,*, pages 25–36. Springer International Publishing, September 1-3 2014. ISBN 978-3-319-10401-0. doi:10.1007/978-3-319-10401-0_3.

Ferretti, Valentina and Pomarico, Silvia. An integrated approach for studying the land suitability for ecological corridors through spatial multicriteria evaluations. *Environment, Development and Sustainability*, 15(3):859–885, 2013.

Figueira, José; Greco, Salvatore, and Ehrgott, Matthias. *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 78. Springer Science+Business Media, 2005.

Filatova, Tatiana; Parker, Dawn, and van der Veen, Anne. Agent-Based Urban Land Markets: Agentś Pricing Behavior, Land Prices and Urban Land Use Change.

*Journal of Artificial Societies and Social Simulation*, 12(1):3, 2009. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/12/1/3.html`.

Fitzpatrick, J Michael and Grefenstette, John J. Genetic algorithms in noisy environments. *Machine Learning*, 3(2-3):101–120, 1988.

Floyd, Myron F; Spengler, John O; Maddock, Jason E; Gobster, Paul H, and Suau, Luis J. Park-based physical activity in diverse communities of two US cities: an observational study. *American Journal of Preventive Medicine*, 34(4):299–305, 2008.

Fogel, David Bruce. *Evolving artificial intelligence*. PhD thesis, University of California at San Diego La Jolla, CA, USA, 1992.

Fogel, Lawrence J; Owens, Alvin J, and Walsh, Michael J. *Artificial Intelligence through simulated evolution.* Jonh Wiley & Sons Inc., New York, NY, USA, 1966.

Fong, CO and Srinivasan, V. The multiregion dynamic capacity expansion problem: an improved heuristic. *Management Science*, 32(9):1140–1152, 1986.

Fonseca, Carlos M and Fleming, Peter J. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In *Proceedings of the 5th International Conference on Genetic Algorithms*, volume 93, pages 416–423. Morgan Kaufmann Publishers Inc., 1993.

Fonseca, Carlos M and Fleming, Peter J. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.

Fontaine, Corentin M. *Residential agents and land use change modelling*. PhD thesis, University of Edinburgh, 2010.

Forsyth, Ann and Mussachio, Laura. *Designing Small Parks: a Manual for Addressing Social and Ecological Concerns*. John Wiley & Sons, October 2005.

Fotheringham, A Stewart and Brunsdon, Chris. Some thoughts on inference in the analysis of spatial data. *International Journal of Geographical Information Science*, 18(11):447–457, July-August 2004. doi:10.1080/13658810410001658065.

Frankhauser, Pierre. Fractal geometry of urban patterns and their morphogenesis. *Discrete Dynamics in Nature and Society*, 2(2):127–145, 1998.

Frankhauser, Pierre. Comparing the morphology of urban patterns in Europe–a fractal approach. *European Cities–Insights on outskirts, Report COST Action*, 10: 79–105, 2004.

Frankhauser, Pierre. From Fractal Urban Pattern Analysis to Fractal Urban Planning Concepts. In *Computational Approaches for Urban Environments*, pages 13–48. Springer, 2015.

Frankhauser, Pierre; Peeters, Dominique, and Thomasô, Isabelle. Where Alonso meets Sierpinski: an urban economic model of a fractal metropolitan area. *Environment and Planning A*, 36:1471–1498, 2004.

Frankhauser, Pierre; Thomas, Isabelle, and Badariotti, Dominique. Comparing the fractality of urban districts: do national processes matter in Europe? *Journal of Geographical Systems*, 14(2):189–208, 2012.

Friesz, Terry L. Transportation network equilibrium, design and aggregation: key developments and research opportunities. *Transportation Research Part A: General*, 19(5):413–427, 1985.

Frieze, Alan M. A cost function property for plant location problems. *Mathematical Programming*, 7(1):245–248, 1974.

Fujishige, Satoru. *Submodular Functions and Optimization*, volume 58. Elsevier, 2005.

Fujita, Masahisa. *Urban Economic Theory: Land Use and City Size*. Cambridge University Press, 1989.

Funahashi, Ken-Ichi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

Furlan, Sueli. *Geografias de São Paulo: a metrópole do século XXI. São Paulo*, chapter Paisagens Sustentáveis: São Paulo e sua Cobertura Vegetal. Contexto, 2004.

Galbraith, Jay R. *Designing Complex Organizations*. Addison-Wesley Longman Publishing Co., Inc., 1973.

Galvao, Roberto Dieguez and Santibanez-Gonzalez, Ernesto del R. A Lagrangean heuristic for the p k-median dynamic location problem. *European Journal of Operational Research*, 58(2):250–262, 1992.

García, AM; Santé, Inés; Boullón, Marcos, and Crecente, Rafael. Calibration of an urban cellular automaton model by using statistical techniques and a genetic algorithm. Application to a small urban settlement of NW Spain. *International Journal of Geographical Information Science*, 27(8):1593–1611, 2013. doi:10.1080/13658816.2012.762454.

Garcia, Laurent and Sabbadin, Régis. Possibilistic influence diagrams. *Frontiers in Artificial Intelligence and Applications*, 141:372–376, 2006.

Gary, Michael R and Johnson, David S. *Computers and Intractability: a Guide to NP-completeness*. W.H. Freeman & Company, 1979.

Gaschler, Andre; Petrick, Ronald PA; Giuliani, Manuel; Rickert, Markus, and Knoll, Alois. KVP: A knowledge of volumes approach to robot task planning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan*, pages 202–208. IEEE, November 2013.

Geertman, Stan and Stillwell, John. Planning support systems: an inventory of current practice. *Computers, Environment and Urban Systems*, 28(4):291–310, 2004.

Getis, Arthur and Getis, Judith. Christaller's central place theory. *Journal of Geography*, 65(5):220–226, 1966.

Gibson, Michael J; Keedwell, Edward C, and Savić, Dragan A. An investigation of the efficient implementation of cellular automata on multi-core CPU and GPU hardware. *Journal of Parallel and Distributed Computing*, 77:11–25, 2015.

Gilbert, Nigel and Troitzsch, Klaus G. *Simulation for the Social Scientist*. Open University Press, 2005. ISBN 0335216005.

Giles-Corti, Billie; Broomhall, Melissa H; Knuiman, Matthew; Collins, Catherine; Douglas, Kate; Ng, Kevin; Lange, Andrea, and Donovan, Robert J. Increasing walking: How important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine*, 28(2, Supplement 2):169 – 176, 2005. ISSN 0749-3797. doi:10.1016/j.amepre.2004.10.018. URL `http://www.sciencedirect.com/science/article/pii/S0749379704002983`.

Gillham, Oliver. *The Limitless City: a Primer on the Urban Sprawl Debate*. Island Press, 2002.

Gimblett, Randy; Daniel, Terry; Cherry, Susan, and Meitner, Michael J. The simulation and visualization of complex human–environment interactions. *Landscape and Urban Planning*, 54(1):63–79, 2001.

Giunta, Anthony A and Watson, Layne T. A comparison of approximation modeling techniques: polynomial versus interpolating models. Technical Report AIAA Paper 98-4758, Virginia Polytechnic Institute and State University; Center for Advanced Vehicles; Blacksburg, VA, United States, 1998.

Glickman, Daniel. Building Cities of Green. In Kollin, C., editor, *National Urban Forest of Conference. American Forests, Washington, DC.*, pages 4–7, 1999.

Godbey, Geoffrey; Graefe, Alan R, and James, Stephen W. *The Benefits of Local Recreation and Park Services, a Nationwide Study of the Perceptions of the American Public*. National Recreation and Park Association Ashburn (VA), 1992.

Goldberg, David E. *Genetic Algorithms in Search, Optimisation, and Machine Learning*. Addison-Wesley Boston, MA, 1989.

Goldberg, David E and Deb, Kalyanmoy. A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, 1:69–93, 1991.

Goldengorin, Boris; Sierksma, Gerard; Tijssen, Gert A, and Tso, Michael. The data-correcting algorithm for the minimization of supermodular functions. *Management Science*, 45(11):1539–1551, 1999.

Golovin, Daniel; Krause, Andreas; Gardner, Beth; Converse, Sarah J., and Morey, Steve. Dynamic Resource Allocation in Conservation Planning. *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, Menlo Park, CA*, pages 1331–1336, 16 August 2011.

Gong, Jianzhou; Liu, Yansui, and Chen, Wenli. Land suitability evaluation for development using a matter-element model: A case study in Zengcheng, Guangzhou, China. *Land Use Policy*, 29(2):464–472, 2012.

González, Pablo Barreira; Gómez-Delgado, Montserrat, and Benavente, Francisco-Aguilera. Vector-based Cellular Automata: exploring new methods of urban growth simulation with cadastral parcels and graph theory. In *4th International Conference on Computers in Urban Planning and Urban Management*, July 2015.

Gray, A. The Intuitive Idea of Distance on a Surface. *Modern differential geometry of curves and surface with mathematica. 2nd ed. Boca Raton: CRC Press*, pages 341–345, 1997.

Greeuw, Sandra CH; van Asselt, Marjolein BA; Grosskurth, Jasper; Storms, CAMH; Rijkens-Klomp, Nicole; Rothman, Dale S; Rotmans, Jan, and Ribeiro, Teresa. *Cloudy Crystal Balls*. International Centre for Integrative Studies (ICIS), November 2000.

Grefenstette, John J. Optimization of control parameters for genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 16(1):122–128, 1986.

Griffin, Arthur F and Stanish, Charles. An agent-based model of prehistoric settlement patterns and political consolidation in the Lake Titicaca Basin of Peru and Bolivia. *Structure and Dynamics*, 2(2):1–47, 2007.

Grimm, Nancy B; Faeth, Stanley H; Golubiewski, Nancy E; Redman, Charles L; Wu, Jianguo; Bai, Xuemei, and Briggs, John M. Global change and the ecology of cities. *Science*, 319(5864):756–760, 2008.

Grimm, Volker and Railsback, Steven F. Designing, formulating, and communicating agent-based models. In *Agent-based Models of Geographical Systems*, pages 361–377. Springer, 2012.

Grimm, Volker and Railsback, Steven F. *Individual-based Modeling and Ecology*. Princeton University Press, 2013.

Groenewegen, Peter P; Berg, Agnes E; Vries, Sjerp, and Verheij, Robert A. Vitamin G: effects of green space on health, well-being, and social safety. *BMC Public Health*, 6(1):1, 2006.

Guezguez, Wided; Amor, Nahla Ben, and Mellouli, Khaled. Qualitative possibilistic influence diagrams based on qualitative possibilistic utilities. *European Journal of Operational Research*, 195(1):223–238, 2009.

Guo, Zhongwei; Xiao, Xiangming, and Li, Dianmo. An assessment of ecosystem services: water flow regulation and hydroelectric power production. *Ecological Applications*, 10(3):925–936, 2000.

Hajela, Prabhat and Lin, C-Y. Genetic search strategies in multicriterion optimal design. *Structural Optimization*, 4(2):99–107, 1992.

Hales, David; Rouchier, Juliette, and Edmonds, Bruce. Model-to-model analysis. *Journal of Artificial Societies and Social Simulation*, 6(4), 2003.

Hall, Troy and Shelby, Bo. Temporal and spatial displacement: evidence from a high-use reservoir and alternate sites. *Journal of Leisure Research*, 32(4):435–457, January 2000.

Hamacher, Horst W and Drezner, Zvi. *Facility Location: Applications and Theory*. Springer Science+Business Media, 2002.

Hamill, Lynne and Gilbert, Nigel. *Agent-based Modelling in Economics*. John Wiley & Sons, 2015.

Hamilton, Bruce W and Röell, Ailsa. Wasteful commuting. *The Journal of Political Economy*, pages 1035–1053, 1982.

Hammel, Ulrich and Bäck, Thomas. Evolution strategies on noisy functions how to improve convergence properties. In *Parallel Problem Solving from Nature—PPSN III*, pages 159–168. Springer, 1994.

Hanski, Ilkka; von Hertzen, Leena; Fyhrquist, Nanna; Koskinen, Kaisa; Torppa, Kaisa; Laatikainen, Tiina; Karisola, Piia; Auvinen, Petri; Paulin, Lars, and Mäkelä, Mika J. Environmental biodiversity, human microbiota, and allergy are interrelated. *Proceedings of the National Academy of Sciences*, 109(21):8334–8339, 2012.

Harik, George; Cantú-Paz, Erick; Goldberg, David E, and Miller, Brad L. The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, 7(3):231–253, 1999.

Hartig, Terry and Staats, Henk. Guest editors' introduction: Restorative environments. *Journal of Environmental Psychology*, 23(2):103–107, 2003.

Hartig, Terry; Mang, Marlis, and Evans, Gary W. Restorative effects of natural environment experiences. *Environment and Behavior*, 23(1):3–26, 1991.

Haviland-Jones, Jeannette; Rosario, Holly Hale; Wilson, Patricia, and McGuire, Terry R. An environmental approach to positive emotion: Flowers. *Evolutionary Psychology*, 3:104–132, 2005.

Health Council of the Netherlands and RMNO, . Nature and the Environment: The influence of nature on social, psychological and physical well-being. *The Hague: Health Council of the Netherlands and RMNO*, 2004.

Heath, Brian; Hill, Raymond, and Ciarallo, Frank. A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9, 2009.

Heimbürger, Peter. Land Policy in Sweden. *Stockholm: United Nations Conference on Human Settlements (Ministry of Housing and Physical Planning)*, 1976.

Herzog, Thomas R; Black, Andrea M; Fountaine, Kimberlee A, and Knotts, Deborah J. Reflection and attentional recovery as distinctive benefits of restorative environments. *Journal of Environmental Psychology*, 17(2):165–170, 1997.

Hilton, Amy B Chan and Culver, Teresa B. Groundwater remediation design under uncertainty using genetic algorithms. *Journal of Water Resources Planning and Management*, 2005.

Hitchings, Russell. Studying the preoccupations that prevent people from going into green space. *Landscape and Urban Planning*, 118:98–102, 2013.

Hokamp, Sascha. Dynamics of tax evasion with back auditing, social norm updating, and public goods provision–An agent-based simulation. *Journal of Economic Psychology*, 40:187–199, 2014.

Holland, John H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The University of Michigan Press, 1975.

Hoover, Edgar M and Giarratani, Frank. An Introduction to Regional Economics, The Web Book of Regional Science, 1984.

Horn, Jeffrey; Nafpliotis, Nicholas, and Goldberg, David E. A niched Pareto genetic algorithm for multiobjective optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence.*, pages 82–87, 1994.

Horner, Mark W and Grubesic, Tony H. A GIS-based planning approach to locating urban rail terminals. *Transportation*, 28(1):55–77, 2001.

Hornik, Kurt. The comprehensive R archive network. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(4):394–398, 2012.

Hornik, Kurt; Stinchcombe, Maxwell, and White, Halbert. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Howard, RA. *Dynamic Programming and Markov Processes*. Technology Press Research Monographs. MIT Press, June 1960.

Huang, Guohe; Baetz, Brian W, and Patry, Gilles G. A grey linear programming approach for municipal solid waste management planning under uncertainty. *Civil Engineering Systems*, 9(4):319–335, 1992.

Huang, Qingxu; Parker, Dawn C; Filatova, Tatiana, and Sun, Shipeng. A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, 40, 2013.

Hughes, Evan J. Evolutionary multi-objective ranking with uncertainty and noise. In *Evolutionary Multi-Criterion Optimization*, pages 329–343. Springer, 2001.

Ibaraki, Toshihide and Katoh, Naoki. *Resource Allocation Problems: Algorithmic Approaches*. The MIT Press, 1988.

Iovine, G; D'Ambrosio, D, and Di Gregorio, S. Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. *Geomorphology*, 66(1):287–303, 2005.

Irwin, Elena G and Bockstael, Nancy E. Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*, 2 (1):31–54, 2002.

Irwin, Elena G; Bockstael, Nancy E, and Cho, Hyun Jin. Measuring and modeling urban sprawl: Data, scale and spatial dependencies. In *Regional Science Association Meetings, Toronto, Canada, November*, 2006.

Itami, Robert M. Simulating spatial dynamics: cellular automata theory. *Landscape and Urban Planning*, 30(1):27–47, 1994.

Iverson, Louis R and Cook, Elizabeth A. Urban forest cover of the Chicago region and its relation to household density and income. *Urban Ecosystems*, 4(2):105–124, 2000.

Jacobs, Jane. *The Death and Life of Great American Cities*. Vintage Books, NY, USA, 1961.

Jansen, Thomas and Wegener, Ingo. The analysis of evolutionary algorithms–A proof that crossover really can help. *Algorithmica*, 34(1):47–66, 2002. ISSN 1432-0541. doi:10.1007/s00453-002-0940-2. URL `http://dx.doi.org/10.1007/s00453-002-0940-2`.

Janssen, Ron and Rietveld, Piet. Multicriteria analysis and geographical information systems: an application to agricultural land use in the Netherlands. In *Geographical Information Systems for Urban and Regional Planning*, pages 129–139. Springer, 1990.

Jantz, Claire A; Goetz, Scott J; Donato, David, and Claggett, Peter. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*, 34(1):1–16, 2010.

Jawarneh, Rana N; Julian, Jason P, and Lookingbill, Todd R. The influence of physiography on historical and future land development changes: A case study of central Arkansas (USA), 1857–2030. *Landscape and Urban Planning*, 143:76–89, 2015.

Jeantet, Gildas; Perny, Patrice, and Spanjaard, Olivier. Sequential Decision Making with Rank Dependent Utility: A Minimax Regret Approach. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 1931–1937. AAAI Press, 2012. URL `http://dl.acm.org/citation.cfm?id=2900929.2901001`.

Jensen, Mikkel T. Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms. *IEEE Transactions on Evolutionary Computation*, 7(5):503–515, 2003.

Jim, CY and Chen, Wendy Y. Perception and attitude of residents toward urban green spaces in Guangzhou (China). *Environmental Management*, 38(3):338–349, 2006.

Jin, Yaochu. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing*, 9(1):3–12, 2005.

Jin, Yaochu and Branke, Jürgen. Evolutionary optimization in uncertain environments-A survey. *IEEE Transactions on Evolutionary Computation*, 9 (3):303–317, 2005.

Jin, Yaochu; Olhofer, Markus, and Sendhoff, Bernhard. A framework for evolutionary optimization with approximate fitness functions. *IEEE Transactions on Evolutionary Computation*, 6(5):481–494, 2002.

Jjumba, Anthony and Dragićević, Suzana. High resolution urban land-use change modeling: agent iCity approach. *Applied Spatial Analysis and Policy*, 5(4):291–315, 2012.

Jo, H-K and McPherson, E Gregory. Indirect carbon reduction by residential vegetation and planting strategies in Chicago, USA. *Journal of Environmental Management*, 61(2):165–177, 2001.

Johnson-Laird, Philip N. Mental models in cognitive science. *Cognitive Science*, 4 (1):71–115, 1980.

Kaczynski, Andrew T; Potwarka, Luke R, and Saelens, Brian E. Association of park size, distance, and features with physical activity in neighborhood parks. *American Journal of Public Health*, 98(8):1451, 2008.

Kapelan, Zoran S; Savic, Dragan A, and Walters, Godfrey A. Multiobjective design of water distribution systems under uncertainty. *Water Resources Research*, 41 (11), 2005.

Kaplan, Rachel. The analysis of perception via preference: a strategy for studying how the environment is experienced. *Landscape Planning*, 12(2):161–176, 1985.

Kaplan, Rachel and Kaplan, Stephen. *The Experience of Nature: A Psychological Perspective*. CUP Archive, 1989.

Kaplan, Stephen. Parks for the future: A psychological perspective. *Stad och Land (Sweden)*, 1990.

Kariv, Oded and Hakimi, S Louis. An algorithmic approach to network location problems. II: The p-medians. *SIAM Journal on Applied Mathematics*, 37(3): 539–560, 1979.

Kawachi, Ichiro and Berkman, Lisa. Social cohesion, social capital, and health. *Social Epidemiology*, pages 174–190, 2000.

Kennedy, Marc; Anderson, Clive; O'Hagan, Anthony; Lomas, Mark; Woodward, Ian; Gosling, John Paul, and Heinemeyer, Andreas. Quantifying uncertainty in the biospheric carbon flux for England and Wales. In *Journal of the Royal Statistical Society A*, volume 171, pages 109–135, 2006.

Kim, Donghan and Batty, Michael. Calibrating Cellular Automata Models for Simulating Urban Growth: Comparative Analysis of SLEUTH and Metronamica. *Centre for Advanced Spatial Analysis. Working Paper Series*, 176, 2011.

Kline, Jeffrey D. Public Demand for Preserving Local Open Space. *Society & Natural Resources*, 19(7):645–659, 2006. doi:10.1080/08941920600742419. URL `http://www.tandfonline.com/doi/abs/10.1080/08941920600742419`.

Knight, Frank H. Diminishing returns from investment. *Journal of Political Economy*, 52(1):26–47, 1944.

Knowles, Joshua and Corne, David. The Pareto Archived Evolution Strategy: A new baseline algorithm for Pareto multiobjective optimisation. In *Proceedings of the 1999 Congress on Evolutionary Computation*, volume 1. IEEE, 1999.

Knowles, Joshua and Corne, David. On metrics for comparing nondominated sets. In *Proceedings of the 2002 Congress on Evolutionary Computation CEC'02*, volume 1, pages 711–716. IEEE, 2002.

Knowles, Joshua D and Corne, David W. Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary computation*, 8(2): 149–172, 2000.

Kocabas, Verda and Dragicevic, Suzana. Enhancing a GIS cellular automata model of land use change: Bayesian networks, influence diagrams and causality. *Transactions in GIS*, 11(5):681–702, 2007.

Kondo, Hiroaki and Liu, Fa-Hua. A study on the urban thermal environment obtained through one-dimensional urban canopy model. *Journal of Japan Society for Atmospheric Environment/Taiki Kankyo Gakkaishi*, 33(3):179–192, 1998.

Kotchen, Matthew J and Powers, Shawn M. Explaining the appearance and success of voter referenda for open-space conservation. *Journal of Environmental Economics and Management*, 52(1):373–390, 2006.

Koza, John R. *Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems*. Stanford University, Department of Computer Science, 1990.

Koza, John R. Concept formation and decision tree induction using the genetic programming paradigm. *Parallel Problem Solving from Nature*, pages 124–128, 1991.

Koza, John R. *Genetic Programming: on the Programming of Computers by Means of Natural Selection*, volume 1. MIT press, 1992.

Krause, Andreas and Guestrin, Carlos E. Near-optimal nonmyopic value of information in graphical models. *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 324–331, 2012.

Krause, Andreas; Golovin, Daniel, and Converse, Sarah. Sequential Decision Making in Computational Sustainability Through Adaptive Submodularity. *AI Magazine*, 35(2):8–18, 2014.

Krause, Eugene F. *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. Courier Corporation, 2012.

Kremen, Claire and Ostfeld, Richard S. A call to ecologists: measuring, analyzing, and managing ecosystem services. *Frontiers in Ecology and the Environment*, 3 (10):540–548, 2005.

Kuby, Michael J. Programming Models for Facility Dispersion: The p-Dispersion and Maxisum Dispersion Problems. *Geographical Analysis*, 19(4):315–329, 1987.

Kuenne, Robert E and Soland, Richard M. Exact and approximate solutions to the multisource Weber problem. *Mathematical Programming*, 3(1):193–209, 1972.

Kuo, Frances E and Sullivan, William C. Environment and crime in the inner city does vegetation reduce crime? *Environment and Behavior*, 33(3):343–367, 2001.

Kuo, Frances E; Bacaicoa, Magdalena, and Sullivan, William C. Transforming inner-city landscapes trees, sense of safety, and preference. *Environment and Behavior*, 30(1):28–59, 1998.

Lafe, Olurinde. *Cellular Automata Transforms: Theory and Applications in Multimedia Compression, Encryption, and Modeling*, volume 16. Springer Science & Business Media, 2012.

Langdon, William B. Quadratic Bloat in Genetic Programming. In *Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation (GECCO)*, volume 2000, pages 451–458, 2000.

Lankford, Samuel V and Howard, Dennis R. Developing a tourism impact attitude scale. *Annals of Tourism Research*, 21(1):121–139, 1994.

Latruffe, Laure and Le Mouël, Chantal. Capitalization of government support in agricultural land prices: What do we know? *Journal of Economic Surveys*, 23(4): 659–691, 2009.

Latruffe, Laure and Minviel, Jean Joseph. Determinants of agricultural land price in Brittany: the role of quotas, environmental and land regulations. In *6th*

*INRA-SFER-CIRAD Conference on Research in Social Sciences, Toulouse, France*, December 2012.

Latruffe, Laure; Minviel, Jean Joseph; Salanié, Julien, and others, . The role of environmental and land transaction regulations on agricultural land price: The example of the French region Brittany. *30th Journées de Microéconomie Appliquée*, pages 6–7, 2013.

Le Quéré, Corinne; Moriarty, Roisin; Andrew, Robbie M; Peters, Glen Philip; Ciais, Philippe; Friedlingstein, Pierre; Jones, Samantha D; Sitch, Stephen; Tans, Pieter; Arneth, Almut, and others, . Global Carbon Budget 2014. *Earth System Science Data*, 7(1):47–85, 2015.

Leao, Simone; Bishop, Ian, and Evans, David. Simulating urban growth in a developing nation's region using a cellular automata-based model. *Journal of Urban Planning and Development*, 130(3):145–158, 2004.

Leather, Phil; Pyrgas, Mike; Beale, Di, and Lawrence, Claire. Windows in the Workplace Sunlight, View, and Occupational Stress. *Environment and Behavior*, 30(6):739–762, 1998.

Lebel, Louis; Garden, Po; Banaticla, Ma; Regina, N; Lasco, Rodel D; Contreras, Antonio; Mitra, AP; Sharma, Chhemendra; Nguyen, Hoang Tri; Ooi, Giok Ling, and others, . Management into the development strategies of urbanizing regions in Asia: implications of urban function, form, and role. *Journal of Industrial Ecology*, 11(2):61–81, 2007.

Lee, Andrew CK and Maheswaran, Ravi. The health benefits of urban green spaces: a review of the evidence. *Journal of Public Health*, 33(2):212–222, 2011.

Lee, Douglass B. Requiem for large-scale models. *Journal of the American Planning Association*, 39:163 – 78, 1973.

Lee, Douglass B. Retrospective on large-scale urban models. *Journal of the American Planning Association*, 60(1):35–40, 1994.

Lee, HY. An analysis on development capacity of an urbanized area for urban growth management. *Journal of the Korean Urban Geographical Society*, 11(1):1–18, 2008.

Lee, Sang-Bum and Luss, Hanan. Multifacility-type capacity expansion planning: algorithms and complexities. *Operations Research*, 35(2):249–253, 1987.

Lehmann, Niklaus; Briner, Simon, and Finger, Robert. The impact of climate and price risks on agricultural land use and crop management decisions. *Land Use Policy*, 35:119–130, 2013.

Lewis, Charles A. *Green Nature/Human Nature: the Meaning of Plants in our Lives*. University of Illinois Press, 1996.

Leyden, Kevin M. Social capital and the built environment: the importance of walkable neighborhoods. *American Journal of Public Health*, 93(9):1546–1551, 2003.

Li, Lin; Sato, Yohei, and Zhu, Haihong. Simulating spatial urban expansion based on a physical process. *Landscape and Urban Planning*, 64(1):67–76, 2003.

Li, Qing; Morimoto, K; Kobayashi, M; Inagaki, H; Katsumata, M; Hirata, Y; Hirata, K; Shimizu, T; Li, YJ; Wakayama, Y, and others, . A forest bathing trip increases human natural killer activity and expression of anti-cancer proteins in female subjects. *Journal of Biological Regulators and Homeostatic Agents*, 22(1):45–55, 2007a.

Li, Qing; Morimoto, K; Kobayashi, M; Inagaki, H; Katsumata, M; Hirata, Yukiyo; Hirata, Kimiko; Suzuki, Hiroko; Li, YJ; Wakayama, Y, and others, . Visiting a forest, but not a city, increases human natural killer activity and expression of anti-cancer proteins. *International Journal of Immunopathology and Pharmacology*, 21(1):117–127, 2007b.

Li, Tong-Hua; Lucasius, Carlos B, and Kateman, Gerrit. Optimization of calibration data with the dynamic genetic algorithm. *Analytica Chimica Acta*, 268(1):123–134, 1992.

Li, Xia and Yeh, Anthony Gar-On. Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environment and Planning A*, 33(8):1445–1462, 2001.

Li, Xia and Yeh, Anthony Gar-On. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16(4):323–343, 2002.

Li, Xia; Lin, Jinyao; Chen, Yimin; Liu, Xiaoping, and Ai, Bin. Calibrating cellular automata based on landscape metrics by using genetic algorithms. *International Journal of Geographical Information Science*, 27(3):594–613, 2013.

Li, YP and Huang, GH. Interval-parameter robust optimization for environmental management under uncertainty. *Canadian Journal of Civil Engineering*, 36(4): 592–606, 2009.

Ligmann-Zielinska, A; Church, RL, and Jankowski, P. Sustainable urban land use allocation with spatial optimization. In *The 8th International Conference on Geocomputation, Michigan, USA*, pages 1–3, August 2005.

Ligmann-Zielinska, Arika and Jankowski, Piotr. Agent-based models as laboratories for spatially explicit planning policies. *Environment and Planning B: Planning and Design*, 34(2):316, 2007.

Ligtenberg, Arend; Bregt, Arnold K, and Van Lammeren, Ron. Multi-actor-based land use modelling: spatial planning using agents. *Landscape and Urban Planning*, 56(1):21–33, 2001.

Lindhagen, Anders. Forest recreation in Sweden. Four case studies using quantitative and qualitative methods. *Rapport-Institutionen för Skoglig Landskapsvård, Sveriges Lantbruksuniversitet*, (64), 1996.

Linke, Sophia Christine. *Local level application of the dynamic land use model METRONAMICA. Assessment and modeling–a case study on the Dutch municipality Weert*. PhD thesis, Diploma Thesis. Available at the Technical University Berlin Institute of Landscape Architecture and Environmental Planning, 2008.

Littke, Hélène. Planning the Green Walkable City: Conceptualizing Values and Conflicts for Urban Green Space Strategies in Stockholm. *Sustainability*, 7(8): 11306–11320, 2015.

Liu, XiaoPing; Li, Xia; Yeh, Anthony Gar-On; He, JinQiang, and Tao, Jia. Discovery of transition rules for geographical cellular automata by using ant colony optimization. *Science in China Series D: Earth Sciences*, 50(10):1578–1588, 2007.

Liu, Yan and Phinn, Stuart R. Modelling urban development with cellular automata incorporating fuzzy-set approaches. *Computers, Environment and Urban Systems*, 27(6):637–658, 2003.

Loibl, W. and Toetzer, T. Modeling growth and densification processes in suburban regions—simulation of landscape transition with spatial agents. *Environmental Modelling & Software*, 18(6):553 – 563, 2003. ISSN 1364-8152. doi:10.1016/S1364-8152(03)00030-6. URL `http://www.sciencedirect.com/science/article/pii/S1364815203000306`.

Longley, Paul A and Mesev, Victor. On the measurement and generalization of urban form. *Environment and Planning A*, 32(3):473–488, 2000.

Loukaitou-Sideris, Anastasia and Stieglitz, Orit. Children in Los Angeles parks: a study of equity, quality and children's satisfaction with neighbourhood parks. *Town Planning Review*, 73(4):467–488, 2002.

Low, Setha; Taplin, Dana, and Scheld, Suzanne. *Rethinking Urban Parks: Public Space and Cultural Diversity*. University of Texas Press, 2009.

Lu, Fadian and Eriksson, Ljusk Ola. Formation of harvest units with genetic algorithms. *Forest Ecology and Management*, 130(1–3):57 – 67, 2000. ISSN 0378-1127. doi:10.1016/S0378-1127(99)00185-1. URL `http://www.sciencedirect.com/science/article/pii/S0378112799001851`.

Lu, Haiming and Yen, Gary G. Rank-density-based multiobjective genetic algorithm and benchmark test function study. *IEEE Transactions on Evolutionary Computation*, 7(4):325–343, 2003.

Luke, Sean and Panait, Liviu. A comparison of bloat control methods for genetic programming. *Evolutionary Computation*, 14(3):309–344, 2006.

Luo, Linbo; Zhou, Suiping; Cai, Wentong; Lees, Michael, and Low, Malcolm Yoke Hean. Modeling human-like decision making for virtual agents in time-critical situations. In *2010 International Conference on Cyberworlds (CW)*, pages 360–367. IEEE, 2010.

Luttik, Joke. The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and Urban Planning*, 48(3):161–167, 2000.

Lutzenhiser, Margot and Netusil, Noelwah R. The effect of open spaces on a home's sale price. *Contemporary Economic Policy*, 19(3):291–298, 2001.

Lv, Ying; Huang, GH; Li, YP; Yang, ZF; Liu, Y, and Cheng, Guanhui. Planning regional water resources system using an interval fuzzy bi-level programming method. *Journal of Environmental Informatics*, 16(2):43–56, 2010.

Ma, Shifa; He, Jianhua; Liu, Feng, and Yu, Yan. Land-use spatial optimization based on PSO algorithm. *Geo-spatial Information Science*, 14(1):54–61, 2011.

Maas, Jolanda; Verheij, Robert A; Groenewegen, Peter P; De Vries, Sjerp, and Spreeuwenberg, Peter. Green space, urbanity, and health: how strong is the relation? *Journal of Epidemiology and Community Health*, 60(7):587–592, 2006.

Maas, Jolanda; Verheij, Robert A; de Vries, Sjerp; Spreeuwenberg, Peter; Schellevis, Francois G, and Groenewegen, Peter P. Morbidity is related to a green living environment. *Journal of Epidemiology and Community Health*, 63(12):967–973, 2009.

Mahan, Brent L; Polasky, Stephen, and Adams, Richard M. Valuing urban wetlands: a property price approach. *Land Economics*, 76(1), 2000.

Mahiny, Abdolrassoul S and Gholamalifard, Mehdi. Dynamic spatial modeling of urban growth through cellular automata in a GIS environment. *International Journal of Environmental Research*, 1(3):272–279, 2007.

Malczewski, Jacek. Central facility location and environmental health. *Environment and Planning A*, 23(3):385–395, 1991.

Malczewski, Jacek. *GIS and Multicriteria Decision Analysis*. John Wiley & Sons, 1999.

Manne, Alan Sussmann. *Investments for Capacity Expansion: Size, Location, and Time-Phasing*, volume 5. The MIT Press, 1967.

Manning, Robert E and Valliere, William A. Coping in outdoor recreation: Causes and consequences of crowding and conflict among community residents. *Journal of Leisure Research*, 33(4):410–426, 2001.

Manson, Steven M. Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, Ecosystems & Environment*, 111(1):47–62, 2005.

Maoh, Hannah and Kanaroglou, Pavlos. A tool for evaluating urban sustainability via integrated transportation and land use simulation models. *Environnement Urbain/Urban Environment*, 3:28–46, 2009.

Marinoni, Oswald; Higgins, A; Hajkowicz, Stefan, and Collins, K. The multiple criteria analysis tool (MCAT): A new software tool to support environmental investment decision making. *Environmental Modelling & Software*, 24(2):153–164, 2009.

Markon, Sandor; Arnold, Dirk V; Back, Thomas; Beielstein, Thomas, and Beyer, Hans-Georg. Thresholding-a selection operator for noisy ES. In *Proceedings of the 2001 Congress on Evolutionary Computation*, volume 1, pages 465–472. IEEE, 2001.

Maruani, Tseira and Amit-Cohen, Irit. Open space planning models: A review of approaches and methods. *Landscape and Urban Planning*, 81(1–2):1 – 13, 2007. ISSN 0169-2046. doi:http://dx.doi.org/10.1016/j.landurbplan.2007.01.003.

Masoomi, Zohreh; Mesgari, Mohammad Sadi, and Hamrah, Majid. Allocation of urban land uses by Multi-Objective Particle Swarm Optimization algorithm. *International Journal of Geographical Information Science*, 27(3):542–566, 2013.

Matthews, Robin B; Gilbert, Nigel G; Roach, Alan; Polhill, J Gary, and Gotts, Nick M. Agent-based land-use models: a review of applications. *Landscape Ecology*, 22(10):1447–1459, 2007.

Mayor of London, . Open Space Strategy: Best Practice Guidance. Technical report, London: Commission for Architecture and the Built Environment and the Greater London Authority, 2009. URL `https://www.designcouncil.org.uk/sites/default/files/asset/document/open-space-strategies.pdf`.

McCormack, Gavin R; Giles-Corti, Billie; Bulsara, Max, and Pikora, Terri J. Correlates of distances traveled to use recreational facilities for physical activity behaviors. *International Journal of Behavioral Nutrition and Physical Activity*, 3 (1):18, 2006.

McCormack, Gavin R; Rock, Melanie; Toohey, Ann M, and Hignell, Danica. Characteristics of urban parks associated with park use and physical activity: a review of qualitative research. *Health & Place*, 16(4):712–726, 2010.

McHale, Melissa R; Gregory McPherson, E, and Burke, Ingrid C. The potential of urban tree plantings to be cost effective in carbon credit markets. *Urban Forestry & Urban Greening*, 6(1):49–60, 2007.

McMillen, Daniel P. Testing for monocentricity. *A Companion to Urban Economics. Oxford: Blackwell*, pages 128–140, 2006.

McPherson, E Gregory. Urban Forestry in North America. *Renewable Resources Journal*, 24(3):8–12, 2006.

McPherson, E Gregory and Simpson, James R. Potential energy savings in buildings by an urban tree planting programme in California. *Urban Forestry & Urban Greening*, 2(2):73–86, 2003.

Melachrinoudis, Emanuel; Min, Hokey, and Wu, Xing. A multiobjective model for the dynamic location of landfills. *Location Science*, 3(3):143–166, 1995.

Melo, M Teresa; Nickel, Stefan, and Da Gama, F Saldanha. Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. *Computers & Operations Research*, 33(1): 181–208, 2006.

Metropolis, Nicholas and Ulam, Stanislaw. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

Michalewicz, Zbigniew. A Survey of Constraint Handling Techniques in Evolutionary Computation Methods. *Evolutionary Programming*, 4:135–155, 1995.

Miettinen, Kaisa; Mäkelä, Marko M, and Toivanen, Jari. Numerical comparison of some penalty-based constraint handling techniques in genetic algorithms. *Journal of Global Optimization*, 27(4):427–446, 2003.

Millenium Ecosystem Assessment, . Millennium ecosystem assessment. *Eco systems and Human Well-Being: A Framework for Assessment. Island Press, Washington, DC*, 2003.

Miller, Brad L. Noise, sampling, and efficient genetic algorithms. *Doctoral Dissertation. University of Illinois at Urbana-Champaign*, 1997.

Miller, Brad L and Goldberg, David E. Optimal sampling for genetic algorithms. *Urbana*, 51:61801, 1996.

Miller, Eric J.; Hunt, John D.; Abraham, John E., and Salvini, Paul A. Microsimulating urban systems. *Computers, Environment and Urban Systems*, 28(1-2):9 – 44, 2004. ISSN 0198-9715. doi:10.1016/S0198-9715(02)00044-3. Geosimulation.

Mills, David E. Growth, speculation and sprawl in a monocentric city. *Journal of Urban Economics*, 10(2):201–226, 1981.

Mills, Edwin S. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, pages 197–210, 1967.

Mills, Edwin S and MacKinnon, James. Notes on the new urban economics. *The Bell Journal of Economics and Management Science*, pages 593–601, 1973.

Ministerio de Medio Ambiente, . *Consumo Sostenible en España. Sustainable Consumption in Spain.* Centro de Publicaciones, Madrid, 2000.

Mirchandani, Pitu B and Francis, Richard L. *Discrete Location Theory.* Wiley-Interscience, 1990.

Mitchell, Melanie. *An Introduction to Genetic Algorithms.* MIT press, 1998.

Mitchell, Richard and Popham, Frank. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650): 1655–1660, 2008.

Mitsova, Diana; Shuster, William, and Wang, Xinhao. A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landscape and Urban Planning*, 99(2):141–153, 2011.

Moghadam, Hossein Shafizadeh and Helbich, Marco. Spatiotemporal urbanization processes in the megacity of Mumbai, India: a Markov chains-cellular automata urban growth model. *Applied Geography*, 40:140–149, 2013.

Molotch, Harvey. The City as a Growth Machine: Toward a Political Economy of Place. *American Journal of Sociology*, 82(2):pp. 309–332, September 1976. URL `http://www.jstor.org/stable/2777096`.

Moore, George C and Revelle, Charles. The hierarchical service location problem. *Management Science*, 28(7):775–780, 1982.

Mordvintsev, Alexander S; Krzhizhanovskaya, Valeria V; Lees, Michael H, and SLees, Peter. Simulation of city evacuation coupled to flood dynamics. In *Pedestrian and Evacuation Dynamics 2012*, pages 485–499. Springer, 2014.

Mousavi, Seyed Mohsen; Niaki, Seyed Taghi Akhavan; Mehdizadeh, Esmaeil, and Tavarroth, Mohammad Reza. The capacitated multi-facility location–allocation problem with probabilistic customer location and demand: two hybrid metaheuristic algorithms. *International Journal of Systems Science*, 44(10):1897–1912, 2013.

Mubasher, Mian Muhammad and Jaffry, Syed Waqar. Incorporation of the Driver's Personality Profile in an Agent Model. *PROMET – Traffic & Transportation*, 2015.

Munro, David A; Holdgate, Martin W, and others, . *Caring for the Earth: a Strategy for Sustainable Living.* International Union for the Conservation of Nature and Natural Resources, 1991.

Murata, Tadahiko and Ishibuchi, Hisao. MOGA: multi-objective genetic algorithms. In *IEEE International Conference on Evolutionary Computation*, volume 1, page 289, 1995.

Murray, Alan T and Church, Richard L. Heuristic solution approaches to operational forest planning problems. *Operations-Research-Spektrum*, 17(2-3):193–203, 1995.

Murray, Alan T and Church, Richard L. Applying simulated annealing to location-planning models. *Journal of Heuristics*, 2(1):31–53, 1996.

Murray-Rust, Dave; Rieser, Verena; Robinson, Derek T; Miličič, Vesna, and Rounsevell, Mark. Agent-based modelling of land use dynamics and residential quality of life for future scenarios. *Environmental Modelling & Software*, 46(0):75 – 89, 2013. ISSN 1364-8152. doi:http://dx.doi.org/10.1016/j.envsoft.2013.02.011. URL `http://www.sciencedirect.com/science/article/pii/S1364815213000510`.

Muth, Richard F. Cities and housing; the spatial pattern of urban residential land use. *University of Chicago Press*, 1969.

Nalle, Darek J; Arthur, Jeffrey L, and Sessions, John. Designing compact and contiguous reserve networks with a hybrid heuristic algorithm. *Forest Science*, 48 (1):59–68, 2002.

Neema, Meher Nigar and Ohgai, Akira. Multi-objective location modeling of urban parks and open spaces: Continuous optimization. *Computers, Environment and Urban Systems*, 34(5):359–376, 2010.

Nelson, Erik; Mendoza, Guillermo; Regetz, James; Polasky, Stephen; Tallis, Heather; Cameron, DRichard; Chan, Kai MA; Daily, Gretchen C; Goldstein, Joshua; Kareiva, Peter M; Lonsdorf, Eric; Naidoo, Robin; Ricketts, Taylor H, and Shaw, MRebecca. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, 7(1):4–11, 2009. ISSN 1540-9309. doi:10.1890/080023. URL `http://dx.doi.org/10.1890/080023`.

Nemhauser, George L and Wolsey, Laurence A. Maximizing submodular set functions: formulations and analysis of algorithms. *North-Holland Mathematics Studies*, 59: 279–301, 1981.

Nemhauser, George L and Wolsey, Leonard A. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.

Neteler, Markus and Mitasova, Helena. *Open Source GIS: a GRASS GIS Approach*, volume 689. Springer Science & Business Media, 2013.

Nicklow, John; Reed, Patrick; Savic, Dragan; Dessalegne, Tibebe; Harrell, Laura; Chan-Hilton, Amy; Karamouz, Mohammad; Minsker, Barbara; Ostfeld, Avi; Singh, Abhishek, and Zechman, Emily. State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*, 136(4):412–432, 2009.

Niemelä, Jari; Saarela, Sanna-Riikka; Söderman, Tarja; Kopperoinen, Leena; Yli-Pelkonen, Vesa; Väre, Seija, and Kotze, D Johan. Using the ecosystem services approach for better planning and conservation of urban green spaces: a Finland case study. *Biodiversity and Conservation*, 19(11):3225–3243, 2010.

Nilsson, Kjell and Randrup, Thomas B. Urban and Peri-Urban Forestry. In *Proceedings of the XI World Forestry Congress on Forest and tree resources*, volume 1, pages 97 – 110, 1997.

North, Michael J; Howe, Thomas R; Collier, Nicholson T, and Vos, Jerry R. The Repast Simphony Runtime System. In Macal, C.M.; North, M.J., and Sallach, D., editors, *Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms. Argonne, Illinois, USA*, October 2005.

North, Michael J; Collier, Nicholson T, and Vos, Jerry R. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(1):1–25, 2006.

Nosetto, Marcelo D; Jobbágy, Estevan G; Brizuela, Alicia B, and Jackson, Robert B. The hydrologic consequences of land cover change in central Argentina. *Agriculture, Ecosystems & Environment*, 154:2–11, 2012.

Nowak, David J and McPherson, E Gregory. Quantifying the impact of trees: the Chicago Urban Forest Climate Project. *Unasylva*, 44(173):39–44, 1993.

Oberkampf, William L; Helton, Jon C; Joslyn, Cliff A; Wojtkiewicz, Steven F, and Ferson, Scott. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering & System Safety*, 85(1):11–19, 2004.

Ogawa, Hideaki and Fujita, Masahisa. Nonmonocentric urban configurations in a two-dimensional space. *Environment and Planning A*, 21(3):363–374, 1989.

Ohlemüller, M. Tabu search for large location-allocation problems. *Journal of the Operational Research Society*, 48(7):745–750, 1997.

O'Madadhain, Joshua; Fisher, Danyel; Smyth, Padhraic; White, Scott, and Boey, Yan-Biao. Analysis and visualization of network data using JUNG. *Journal of Statistical Software*, 10(2):1–35, 2005.

Ong, Yew S; Nair, Prasanth B, and Keane, Andrew J. Evolutionary optimization of computationally expensive problems via surrogate modeling. *American Institute of Aeronautics and Astronautics (AIAA) Journal*, 41(4):687–696, 2003.

O'Sullivan, David. Graph-cellular automata: a generalised discrete urban and regional model. *Environment and Planning B*, 28(5):687–706, 2001a.

O'Sullivan, David. Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis*, 33(1):1–18, 2001b.

O'Sullivan, David. Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, 4(3):251–274, 2002.

O'Sullivan, David. Changing Neighborhoods—Neighborhoods Changing A Framework for Spatially Explicit Agent-Based Models of Social Systems. *Sociological Methods & Research*, 37(4):498–530, 2009.

Otter, Henriëtte S; van der Veen, Anne, and de Vriend, Huib J. ABLOoM: Location Behaviour, Spatial Patterns, and Agent-Based Modelling. *Journal of Artificial Societies and Social Simulation*, 4, 2001. URL `http://ideas.repec.org/a/jas/jasssj/2001-13-1.html`.

Owen, Susan Hesse and Daskin, Mark S. Strategic facility location: A review. *European Journal of Operational Research*, 111(3):423–447, 1998.

Oyman, A Irfan; Beyer, Hans-georg, and Schwefel, Hans-paul. Analysis of the (1, $\lambda$)-ES on the Parabolic Ridge. *Evolutionary Computation*, 8(3):249–265, 2000.

Palahí, Marc; Pukkala, Timo; Pascual, L, and Trasobares, Antoni. Examining alternative landscape metrics in ecological forest planning: a case for capercaillie in Catalonia. *Forest Systems*, 13(3):527–538, 2004.

Park, Bum-Jin; Tsunetsugu, Yuko; Kasetani, Tamami; Morikawa, Takeshi; Kagawa, Takahide, and Miyazaki, Yoshifumi. Physiological effects of forest recreation in a young conifer forest in Hinokage Town, Japan. *Silva Fenn*, 43(2):291–301, 2009.

Park, Robert E; Burgess, Ernest W, and McKenzie, Roderick Duncan. *The City*. University of Chicago Press, 1925.

Park, Seong-Hyun and Mattson, Richard H. Effects of flowering and foliage plants in hospital rooms on patients recovering from abdominal surgery. *HortTechnology*, 18(4):563–568, 2008.

Park, Soyoung; Jeon, Seongwoo; Kim, Shinyup, and Choi, Chuluong. Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea. *Landscape and Urban Planning*, 99(2):104–114, 2011.

Parker, Dawn C. and Filatova, Tatiana. A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, 32(6):454 – 463, 2008. ISSN 0198-9715. doi:10.1016/j.compenvurbsys.2008.09.012.

Parker, Dawn C. and Meretsky, Vicky. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems & Environment*, 101(2):233–250, 2004.

Parker, Dawn C; Manson, Steven M; Janssen, Marco A; Hoffmann, Matthew J, and Deadman, Peter. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93(2): 314–337, 2003. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.92.2240`.

Parry, Hazel R; Evans, Andrew J, and Morgan, Derek. Aphid population response to agricultural landscape change: a spatially explicit, individual-based model. *Ecological Modelling*, 199(4):451–463, 2006.

Passow, Shirley S. Land Reserves and Teamwork In Planning Stockholm. *Journal of the American Institute of Planners*, 36(3):179–188, 1970. doi:10.1080/01944367008977304. URL `http://www.tandfonline.com/doi/abs/10.1080/01944367008977304`.

Payne, Laura L; Mowen, Andrew J, and Orsega-Smith, Elizabeth. An examination of park preferences and behaviors among urban residents: the role of residential location, race, and age. *Leisure Sciences*, 24(2):181–198, 2002.

Payne, Laura L; Orsega-Smith, Elizabeth; Roy, Mark, and Godbey, Geoffrey C. Local Park Use and Personal Health Among Older Adults: An Exploratory Study. *Journal of Park and Recreation Administration*, 23(2), 2005.

Pham, Thi-Thanh-Hien; Apparicio, Philippe; Séguin, Anne-Marie; Landry, Shawn, and Gagnon, Martin. Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? *Landscape and Urban Planning*, 107(3): 214–224, 2012.

Pickett, Steward TA; Cadenasso, ML; Grove, JM; Nilon, CH; Pouyat, RV; Zipperer, WC, and Costanza, R. Urban ecological systems: linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas. *Annual Review of Ecology and Systematics*, pages 127–157, 2001.

Pineau, Joelle; Gordon, Geoffrey, and Thrun, Sebastian. Anytime point-based approximations for large POMDPs. *Journal of Artificial Intelligence Research*, 27: 335–380, 2006.

Pipattanasomporn, Manisa; Feroze, Hassan, and Rahman, S. Multi-agent systems in a distributed smart grid: Design and implementation. In *IEEE/PES Power Systems Conference and Exposition. PSCE'09*, pages 1–8. IEEE, 2009.

Pitt, Jeremy and Mamdani, Abe. A protocol-based semantics for an agent communication language. In *International Joint Conference on Artificial Intelligence (IJCAI)*, volume 99, pages 486–491, 1999.

Pitzl, Gerald R. *Encyclopedia of Human Geography.* Greenwood Publishing Group, 2004.

Plantinga, Andrew J and Miller, Douglas J. Agricultural land values and the value of rights to future land development. *Land Economics*, 77(1):56–67, 2001.

Plantinga, Andrew J; Lubowski, Ruben N, and Stavins, Robert N. The effects of potential land development on agricultural land prices. *Journal of Urban Economics*, 52(3):561 – 581, 2002. ISSN 0094-1190. doi:10.1016/S0094-1190(02)00503-X. URL `http://www.sciencedirect.com/science/article/pii/S009411900200503X`.

Polat, Zöhre and Yilmaz, Hasan. Park Networks in Built Environment: Suggestions for the Urban Landscape in Turkey. *World Applied Sciences Journal*, 24(11), 2013.

Pontius, Robert Gilmore; Huffaker, Diana, and Denman, Kevin. Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, 179 (4):445–461, 2004.

Porta, Juan; Parapar, Jorge; Doallo, Ramón; Rivera, Francisco F; Santé, Inés, and Crecente, Rafael. High performance genetic algorithm for land use planning. *Computers, Environment and Urban Systems*, 37:45–58, 2013.

Portugali, Juval and Benenson, Itzhak. Artificial planning experience by means of a heuristic cell-space model: simulating international migration in the urban process. *Environment and Planning A*, 27(10):1647–1665, 1995.

Pouta, Eija and Heikkilä, Marjo. *Virkistysalueiden suunnittelu ja hoito*. Ympäristöministeriö, alueidenkäytön osasto, 1998.

Power, Conrad; Simms, Alvin, and White, Roger. Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, 15(1):77–100, 2001.

Pretty, Jules; Peacock, Jo; Sellens, Martin, and Griffin, Murray. The mental and physical health outcomes of green exercise. *International Journal of Environmental Health Research*, 15(5):319–337, 2005.

Pukkala, Timo and Kurttila, Mikko. Examining the performance of six heuristic optimisation techniques in different forest planning problems. *Silva Fennica*, 39 (1):67–80, August 2005.

Pukkala, Timo; Carneiro, Lorenzo Pascual; Trasobares, Antoni, and Lozano, Marc Palahí. Examinig alternative landscape metrics in ecological forest planning: a case for capercaillie in Catalonia. *Investigación agraria. Sistemas y Recursos Forestales*, 13(3):527–538, 2004.

Puterman, Martin L. Markov decision processes: Discrete dynamic stochastic programming. *John Wiley, New York, NY*, 10, 1994. doi:9780470316887.

Qian, Chao; Yu, Yang, and Zhou, Zhi-Hua. Analyzing Evolutionary Optimization in Noisy Environments. *Evolutionary Computation*, 2013.

Qin, Xiaosheng; Huang, Guohe, and Liu, Lei. A Genetic-Algorithm-Aided Stochastic Optimization Model for Regional Air Quality Management under Uncertainty. *Journal of the Air & Waste Management Association*, 60(1):63–71, 2010. doi:10.3155/1047-3289.60.1.63. URL `http://www.tandfonline.com/doi/abs/10.3155/1047-3289.60.1.63`.

Quaas, Martin F. Environmental amenities and public open space in an efficient urban spatial structure. February 2007.

Rada-Vilela, Juan; Johnston, Mark, and Zhang, Mengjie. Population statistics for particle swarm optimization: Resampling methods in noisy optimization problems. *Swarm and Evolutionary Computation*, 17:37–59, 2014.

Rakas, Jasenka; Teodorović, Dušan, and Kim, Taehyung. Multi-objective modeling for determining location of undesirable facilities. *Transportation Research Part D: Transport and Environment*, 9(2):125–138, 2004.

Ramsay, James O.; Ten Berge, Josephus M.F., and Styan, George P.H. Matrix correlation. *Psychometrika*, 49(3):403–423, 1984.

Rastrigin, LA. Extremal control systems. *Theoretical Foundations of Engineering Cybernetics Series*, 3, 1974.

Ratle, Alain. Accelerating the convergence of evolutionary algorithms by fitness landscape approximation. In *Parallel Problem Solving from Nature—PPSN V,Amsterdam, The Netherlands*, pages 87–96. Springer, September 1998.

Ratle, Alain. Optimal sampling strategies for learning a fitness model. In *Proceedings of the 1999 Congress on Evolutionary Computation*, volume 3, 1999.

Rechenberg, Ingo. *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog, Sturgart, Germany, 1973.

Remmel, Tarmo K and Csillag, Ferenc. When are two landscape pattern indices significantly different? *Journal of Geographical Systems*, 5(4):331–351, 2003.

Revelle, Charles S and Eiselt, Horst A. Location analysis: A synthesis and survey. *European Journal of Operational Research*, 165(1):1–19, 2005.

Revelle, Charles S; Eiselt, Horst A, and Daskin, Mark S. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184(3):817–848, 2008.

Ricardo, David. *Principles of Political Economy and Taxation.* George Bell & Sons, 1891.

Richiardi, Matteo G; Leombruni, Roberto; Saam, Nicole J, and Sonnessa, Michele. A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation*, 9(1), 2006.

Riera, Pere. *Rentabilidad Social de las Infraestructuras: las Rondas de Barcelona: un Análisis Coste-beneficio.* Civitas, S.A., 1993.

Rieser, Verena and Lemon, Oliver. *Reinforcement Learning for Adaptive Dialogue Systems: A Data-driven Methodology for Dialogue Management and Natural Language Generation.* Theory and Applications of Natural Language Processing. Springer, 2011.

Rieser, Verena; Robinson, Derek T; Murray-Rust, Dave, and Rounsevell, Mark. A Comparison of Genetic Algorithms and Reinforcement Learning for Optimising Sustainable Forest Management. In *Proceedings of the 11th International Conference on GeoComputation. London, UK*, July 2011.

Riitters, Kurt H; O'neill, RV; Hunsaker, CT; Wickham, James D; Yankee, DH;

Timmins, SP; Jones, KB, and Jackson, BL. A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, 10(1):23–39, 1995.

Riley, C. Comments on Mills and Evans. *Proceedings of Seminar on Land Use Regulation, Lincoln Institute for Land Policy. Cambridge Mass*, July 2002.

Robinson, Derek T and Brown, Daniel G. Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based GIS approach. *International Journal of Geographical Information Science*, 23(9):1211–1232, 2009.

Robinson, Derek T; Murray-Rust, David; Rieser, Verena; Milicic, Vesna, and Rounsevell, M. Modelling the impacts of land system dynamics on human well-being: Using an agent-based approach to cope with data limitations in Koper, Slovenia. *Computers, Environment and Urban Systems. Special Issue: Geoinformatics 2010*, 36(2):164–176, March 2012.

Romanos, Michael C. *Residential Spatial Structure*. Heath, 1976.

Romanos, Michael C. Household location in a linear multi-center metropolitan area. *Regional Science and Urban Economics*, 7(3):233–250, 1977.

Ronald, Nicole; Dignum, Virginia; Jonker, Catholijn; Arentze, Theo, and Timmermans, Harry. On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology*, 54(6):625–638, 2012.

Roodman, Gary M and Schwarz, Leroy B. Extensions of the multi-period facility phase-out model: New procedures and application to a phase-in/phase-out problem. *AIIE Transactions (American Institute of Industrial Engineers)*, 9(1):103–107, 1977.

Roovers, Pieter; Hermy, Martin, and Gulinck, Hubert. Visitor profile, perceptions and expectations in forests from a gradient of increasing urbanisation in central Belgium. *Landscape and Urban Planning*, 59(3):129–145, 2002.

Ross, Stéphane; Pineau, Joelle; Paquet, Sébastien, and Chaib-Draa, Brahim. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*, 32: 663–704, 2008.

Rounsevell, Mark D.A. and Metzger, Marc J. Developing qualitative scenario storylines for environmental change assessment. *Wiley Interdisciplinary Reviews: Climate Change*, 1(4):606–619, 2010. ISSN 1757-7799. doi:10.1002/wcc.63. URL `http://dx.doi.org/10.1002/wcc.63`.

Rounsevell, Mark D.A.; Robinson, Derek T., and Murray-Rust, Dave. From actors to agents in socio-ecological systems models. In *Royal Philosophical Society B*, volume 367, pages 259–269, 2012.

Saaty, Thomas L. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1):83–98, 2008.

Sabbadin, Régis. A possibilistic model for qualitative sequential decision problems under uncertainty in partially observable environments. In Laskey, Kathryn and Prade, Henri, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden*, pages 567–574. Morgan Kaufmann Publishers Inc., Morgan Kaufmann, San Francisco, CA, July 1999.

Şahin, Güvenç and Süral, Haldun. A review of hierarchical facility location models. *Computers & Operations Research*, 34(8):2310–2331, 2007.

Saito, Ikuo; Ishihara, Osamu, and Katayama, Tadahisa. Study of the effect of green areas on the thermal environment in an urban area. *Energy and Buildings*, 15(3): 493–498, 1991.

Sakoda, James M. The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, 1(1):119–132, 1971.

Salvini, Paul and Miller, Eric J. ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2): 217–234, 2005.

Sanders, Lena; Pumain, Denise; Mathian, Hélene; Guérin-Pace, France, and Bura, Stephane. SIMPOP: a multiagent system for the study of urbanism. *Environment and Planning B: Planning and Design*, 24(2):287–305, 1997.

Sanesi, Giovanni and Chiarello, Francesco. Residents and urban green spaces: the case of Bari. *Urban Forestry & Urban Greening*, 4(3):125–134, 2006.

Sano, Yasuhito and Kita, Hajime. Optimization of noisy fitness functions by means of genetic algorithms using history of search with test of estimation. In *Proceedings of the 2002 Congress on Evolutionary Computation, CEC'02*, volume 1, pages 360–365, 2002.

Santé, Inés; García, Andrés M; Miranda, David, and Crecente, Rafael. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape and Urban Planning*, 96(2):108–122, 2010.

Sasaki, Yuya and Box, Paul. Agent-Based Verification of Von Thünen's Location Theory. *Journal of Artificial Societies and Social Simulation*, 6, 2003.

Savas, Emanuel S. On equity in providing public services. *Management Science*, 24 (8):800–808, 1978.

Schaffer, J David. Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms, Pittsburgh, PA, USA, July 1985*, pages 93–100, 1985.

Schelling, Thomas C. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186, 1971.

Schieritz, Nadine and Milling, Peter M. Modeling the forest or modeling the trees. In *Proceedings of the 21st International Conference of the System Dynamics Society*, pages 20–24, 2003.

Schrijver, Alexander. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

Schroeder, Herbert W and Daniel, T. Preferred features of urban parks and forests. *Journal of Arboriculture*, 8(12):317–322, 1982.

Scott, Allen J. Dynamic location-allocation systems: some basic planning strategies. *Environment and Planning*, 3(1):73–82, 1971.

Scottish Natural Heritage, . Health Impact Assessment of Greenspace: A Guide. *Greenspace Scotland*, 2008.

Sefair, Jorge A; Molano, Adriana; Medaglia, Andrés L, and Sarmiento, Olga L. Locating neighborhood parks with a lexicographic multiobjective optimization method. In *Community-Based Operations Research*, pages 143–171. Springer, 2012.

Semboloni, Ferdinando. The dynamic of an urban cellular automata in a 3-D spatial pattern. In *XXI National Conference Aisre: Regional and Urban Growth in a Global Market, Palermo*, 2000a.

Semboloni, Ferdinando. The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B*, 27(4):549–564, 2000b.

Semboloni, Ferdinando. Optimization and control of the urban spatial dynamics. *Complexus*, 2(3-4):204–216, 2004.

Sfa, Fatima Ezahra; Nemiche, Mohamed, and Lopez, Rafael Pla. A Theoretical Learning Model Combining Stochastic Cellular Automata and Economic Indicators to Simulate Land Use Change. *International Journal of Applied Evolutionary Computation (IJAEC)*, 6(3):1–8, 2015.

Shahid, Rizwan; Bertazzon, Stefania; Knudtson, Merril L, and Ghali, William A. Comparison of distance measures in spatial analytical modeling for health service planning. *BMC Health Services Research*, 9(1):200, 2009.

Shan, Jie; Alkheder, Sharaf, and Wang, Jun. Genetic algorithms for the calibration of cellular automata urban growth modeling. *Photogrammetric Engineering & Remote Sensing*, 74(10):1267–1277, 2008.

Shariff, SS Radiah; Moin, Noor Hasnah, and Omar, Mohd. Location allocation modeling for healthcare facility planning in Malaysia. *Computers & Industrial Engineering*, 62(4):1000–1010, 2012.

Shelby, Bo; Bregenzer, N Stewart, and Johnson, Rebecca. Displacement and product shift: empirical evidence from Oregon rivers. *Journal of Leisure Research*, 20(4): 274–288, 1988.

Shi, Wenzhong and Pang, Matthew Yick Cheung. Development of Voronoi-based cellular automata-an integrated dynamic model for Geographical Information Systems. *International Journal of Geographical Information Science*, 14(5):455–474, 2000.

Shi, Yue Jin; Phipps, Timothy T, and Colyer, Dale. Agricultural land values under urbanizing influences. *Land Economics*, pages 90–100, 1997.

Shiha, Amr N and Chavas, Jean-Paul. Capital market segmentation and US farm real estate pricing. *American Journal of Agricultural Economics*, 77(2):397–407, 1995.

Shiyuan, Hu and Deren, Li. Vector cellular automata based geographical entity. *Geoinformatics*, 2004.

Silberstein, Jane and Maser, Chris. *Land-Use Planning for Sustainable Development*. CRC Press, 2013.

Silva, Elisabete A and Clarke, Keith C. Complexity, emergence and cellular urban models: lessons learned from applying SLEUTH to two Portuguese metropolitan areas. *European Planning Studies*, 13(1):93–115, 2005.

Simpson, Timothy W. Comparative studies of metamodeling techniques under multiple modeling criteria. In *8th Symposium on Multidisciplinary Analysis and Optimization*. American Institute of Aeronautics and Astronautics, 2000.

Smalley, J Bryan; Minsker, Barbara S, and Goldberg, David E. Risk-based in situ bioremediation design using a noisy genetic algorithm. *Water Resources Research*, 36(10):3043–3052, 2000.

Smallwood, Richard D and Sondik, Edward J. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5): 1071–1088, 1973.

Smith, V Kerry and Osborne, Laura L. Do contingent valuation estimates pass a "scope" test? A meta-analysis. *Journal of Environmental Economics and Management*, 31(3):287–301, 1996.

Smith, V Kerry; Poulos, Christine, and Kim, Hyun. Treating open space as an urban amenity. *Resource and Energy Economics*, 24(1):107–129, 2002.

Snyder, Lawrence V. Facility location under uncertainty: a review. *IIE Transactions*, 38(7):547–564, 2006.

Sop Shin, Won. The influence of forest view through a window on job satisfaction and job stress. *Scandinavian Journal of Forest Research*, 22(3):248–253, 2007.

Sridharan, Ramaswami. The capacitated plant location problem. *European Journal of Operational Research*, 87(2):203–213, 1995.

Srinivas, Nidamarthi and Deb, Kalyanmoy. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3): 221–248, 1994.

Stevens, Daniel and Dragićević, Suzana. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design*, 34 (4):708–724, 2007.

Stevens, Daniel; Dragicevic, Suzana, and Rothley, Kristina. iCity: A GIS–CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22(6):761–773, 2007.

Stewart, Theodor J; Janssen, Ron, and van Herwijnen, Marjan. A genetic algorithm approach to multiobjective land use planning. *Computers & Operations Research*, 31(14):2293–2313, 2004.

Su, Daniel Z. GIS-based urban modelling: practices, problems, and prospects. *International Journal of Geographical Information Science*, 12(7):651–671, 1998.

Syberfeldt, Anna; Ng, Amos; John, Robert I, and Moore, Philip. Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling. *European Journal of Operational Research*, 204(3):533–544, 2010.

Taha, Haider G; Akbari, Hashem, and Rosenfeld, Arthur H. *Vegetation Canopy Micro-climate: A Field-project in Davis, California*. Applied Science Division, Lawrence Berkeley Laboratory, University of California-Berkeley, 1989.

Tajibaeva, Liaila; Haight, Robert G, and Polasky, Stephen. A discrete-space urban model with environmental amenities. *Resource and Energy Economics*, 30(2): 170–196, 2008.

Takano, Takehito; Nakamura, Keiko, and Watanabe, Masafumi. Urban residential environments and senior citizens' longevity in megacity areas: the importance of walkable green spaces. *Journal of Epidemiology and Community Health*, 56(12): 913–918, 2002.

Takayama, Norimasa; Korpela, Kalevi; Lee, Juyoung; Morikawa, Takeshi; Tsunetsugu, Yuko; Park, Bum-Jin; Li, Qing; Tyrväinen, Liisa; Miyazaki, Yoshifumi, and Kagawa, Takahide. Emotional, Restorative and Vitalizing Effects of Forest and Urban Environments at Four Sites in Japan. *International Journal of Environmental Research and Public Health*, 11(7):7207–7230, 2014.

Takeyama, Masanao and Couclelis, Helen. Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science*, 11(1):73–91, 1997.

Tan, Kay Chen and Goh, Chi Keong. Handling uncertainties in evolutionary multi-objective optimization. In *Computational Intelligence: Research Frontiers*, pages 262–292. Springer, 2008.

Tang, Junmei; Wang, Jiao, and Yao, Zhijun. Spatio-temporal urban landscape change analysis using the Markov chain model and a modified Genetic Algorithm. *International Journal of Remote Sensing*, 28(15):3255–3271, 2007.

Tang, Wenwu. Simulating complex adaptive geographic systems: A geographically aware intelligent agent approach. *Cartography and Geographic Information Science*, 35(4):239–263, 2008.

Tansel, Barbaros C; Francis, Richard L, and Lowe, Timothy J. State of the art—location on networks: a survey. Part II: exploiting tree network structure. *Management Science*, 29(4):498–511, 1983.

Tarzia, Valentina. European Common Indicators: towards a local sustainability profile. *Ambiente Italia Research Institute, Milano, Italy*, 2003.

Tattoni, Clara; Ciolli, Marco, and Ferretti, Fabrizio. The fate of priority areas for conservation in protected areas: a fine-scale Markov chain approach. *Environmental Management*, 47(2):263–278, 2011.

Teich, Jürgen. Pareto-front exploration with uncertain objectives. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 314–328. Springer, 2001.

Thorsnes, Paul. The Value of a Suburban Forest Preserve: Estimates from Sales of Vacant Residential Building Lots. *Land Economics*, 78(3):426–441, August 1 2002. doi:10.3368/le.78.3.426.

Throsby, David. Culture, economics and sustainability. *Journal of Cultural Economics*, 19(3):199–206, 1995.

Tian, Guangjin; Ouyang, Yun; Quan, Quan, and Wu, Jianguo. Simulating spatiotemporal dynamics of urbanization with multi-agent systems—A case study of the Phoenix metropolitan region, USA. *Ecological Modelling*, 222(5):1129–1138, 2011.

Tobler, W. R. *Theory and Decision Library*, chapter Cellular Geography, pages 379–386. Springer Netherlands, Dordrecht, 1979. ISBN 978-94-009-9394-5. doi:10.1007/978-94-009-9394-5_18. URL http://dx.doi.org/10.1007/978-94-009-9394-5_18.

Todaro, Michael P. A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, pages 138–148, 1969.

Tolley, George S. The welfare economics of city bigness. *Journal of Urban Economics*, 1(3):324–345, 1974.

Toregas, Constantine; Swain, Ralph; ReVelle, Charles, and Bergman, Lawrence. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.

Torrens, Paul M. How cellular models of urban systems work (1. Theory). *Centre for Advanced Spatial Analysis. Working Paper Series*, pages 1–68, November 2000.

Torrens, Paul M and Benenson, Itzhak. Geographic automata systems. *International Journal of Geographical Information Science*, 19(4):385–412, 2005.

Tušar, Tea and Filipič, Bogdan. Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. *IEEE Transactions on Evolutionary Computation*, 19(2):225–245, 2015.

Tyrväinen, Liisa. The amenity value of the urban forest: an application of the hedonic pricing method. *Landscape and Urban Planning*, 37(3–4):211 – 222, 1997. ISSN 0169-2046. doi:10.1016/S0169-2046(97)80005-9. URL http://www.sciencedirect.com/science/article/pii/S0169204697800059.

Tyrväinen, Liisa and Miettinen, Antti. Property Prices and Urban Forest Amenities. *Journal of Environmental Economics and Management*, 39(2):205 – 223, 2000.

ISSN 0095-0696. doi:10.1006/jeem.1999.1097. URL `http://www.sciencedirect.com/science/article/pii/S0095069699910975`.

Tyrväinen, Liisa and Väänänen, Hannu. The economic value of urban forest amenities: an application of the contingent valuation method. *Landscape and Urban Planning*, 43(1):105–118, 1998.

Ullman, Edward. A theory of location for cities. *American Journal of Sociology*, pages 853–864, 1941.

Ulrich, Roger S. Natural versus urban scenes some psychophysiological effects. *Environment and Behavior*, 13(5):523–556, 1981.

Ulrich, Roger S; Simons, Robert F; Losito, Barbara D; Fiorito, Evelyn; Miles, Mark A, and Zelson, Michael. Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, 11(3):201–230, 1991.

Uy, Pham Duc and Nakagoshi, Nobukazu. Application of land suitability analysis and landscape ecology to urban greenspace planning in Hanoi, Vietnam. *Urban Forestry & Urban Greening*, 7(1):25–40, 2008.

Vallejo, Marta; Corne, David W., and Rieser, Verena. Evolving Urbanisation Policies - Using a Statistical Model to Accelerate Optimisation over Agent-based Simulations. In *ICAART 2013 - 5th Conference on Agents and Artificial Intelligence. Barcelona, Spain*, pages 171–181. Springer Berlin Heidelberg, February 2013.

Vallejo, Marta; Corne, David W., and Rieser, Verena. Genetic Algorithm evaluation of green search allocation policies in multilevel complex urban scenarios. *Journal of Computational Science JOCS*, 9:57–63, 2015.

van Delden, Hedwig; Escudero, Juan Carlos; Uljee, Inge, and Engelen, Guy. METRONAMICA: A dynamic spatial land use model applied to Vitoria-Gasteiz. In *Virtual Seminar of the MILES Project. Centro de Estudios Ambientales, Vitoria-Gasteiz*, 2005.

Van Gestel, Tony; Suykens, Johan AK; Baestaens, D-E; Lambrechts, Annemie; Lanckriet, Gert; Vandaele, Bruno; De Moor, Bart, and Vandewalle, Joos. Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks*, 12(4):809–821, 2001.

Van Roy, Tony J and Erlenkotter, Donald. A dual-based procedure for dynamic facility location. *Management Science*, 28(10):1091–1105, 1982.

Van Veldhuizen, David A and Lamont, Gary B. Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary Computation*, 8(2):125–147, 2000.

van Vliet, Jasper; White, Roger, and Dragicevic, Suzana. Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems*, 33(1):35–43, 2009.

Vargas, Patricia A; Di Paolo, Ezequiel A; Harvey, Inman, and Husbands, Phil. *The Horizons of Evolutionary Robotics*. The MIT Press, 2014.

Vaske, Jerry J; Donnelly, Maureen P, and Heberlein, Thomas A. Perceptions of crowding and resource quality by early and more recent visitors. *Leisure Sciences*, 3(4):367–381, 1980.

Vatn, Arild and Bromley, Daniel W. Choices without prices without apologies. *Journal of Environmental Economics and Management*, 26(2):129–148, 1994.

Venema, Henry D.; Calamai, Paul H., and Fieguth, Paul. Forest structure optimization using evolutionary programming and landscape ecology metrics. *European Journal of Operational Research*, 164(2):423 – 439, 2005. ISSN 0377-2217. doi:10.1016/j.ejor.2003.11.011.

Verburg, Peter H; de Nijs, Ton CM; van Eck, Jan Ritsema; Visser, Hans, and de Jong, Kor. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28(6):667–690, 2004.

Villegas, Juan G; Palacios, Fernando, and Medaglia, Andrés L. Solution methods for the bi-objective (cost-coverage) unconstrained facility location problem with an illustrative example. *Annals of Operations Research*, 147(1):109–141, 2006.

Von Neumann, John. Theory of self-reproducing automata. *University of Illinois Press, Urbana IL*, 1966.

Von Thünen, Johann Heinrich. *The Isolated State: an English Edition of Der Isolierte Staat*. Pergamon Press, 1966.

Waddell, Paul; Bhat, Chandra; Eluru, Naveen; Wang, Liming, and Pendyala, Ram. Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board*, 2015.

Wagner, Daniel F. Cellular automata and geographic information systems. *Environment and Planning B*, 24:219–234, 1997.

Wahyudi, Agung and Liu, Yan. Cellular automata for urban growth modeling: a chronological review on transition rules. In *13rd International Conference on Computers in Urban Planning and Urban Management*. CUPUM, 2013.

Wakunda, Jürgen and Zell, Andreas. Median-selection for parallel steady-state evolution strategies. In *Parallel Problem Solving from Nature PPSN VI*, pages 405–414. Springer, 2000.

Wan, Xiaotao; Pekny, Joseph F, and Reklaitis, Gintaras V. Simulation-based optimization with surrogate models—application to supply chain management. *Computers & Chemical Engineering*, 29(6):1317–1328, 2005.

Wang, Na-Fu and Yang, Ya-Wen. Target geometry matching problem for hybrid genetic algorithm used to design structures subjected to uncertainty. In *IEEE Congress on Evolutionary Computation. CEC'09*, pages 1644 – 1651, May 2009. doi:10.1109/CEC.2009.4983139.

Wang, Nianfeng F; Zhang, XM, and Yang, YW. A hybrid genetic algorithm for constrained multi-objective optimization under uncertainty and target matching problems. *Applied Soft Computing*, 13(8):3636–3645, 2013. ISSN 1568-4946. doi:10.1016/j.asoc.2013.03.013. URL `http://www.sciencedirect.com/science/article/pii/S1568494613001208`.

Wang, Qian; Batta, Rajan; Bhadury, Joyendu, and Rump, Christopher M. Budget constrained location problem with opening and closing of facilities. *Computers & Operations Research*, 30(13):2047–2069, 2003.

Wang, Yang and Li, Shuangcheng. Simulating multiple class urban land-use/cover changes by RBFN-based CA model. *Computers & Geosciences*, 37(2):111–121, 2011.

Ward, Douglas P; Murray, Alan T, and Phinn, Stuart R. A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems*, 24 (6):539–558, 2000.

Ward Thompson, Catharine. Urban open space in the 21st century. *Landscape and Urban Planning*, 60(2):59–72, 2002.

Ward Thompson, Catharine and Aspinall, Peter A. Natural environments and their impact on activity, health, and quality of life. *Applied Psychology: Health and Well-Being*, 3(3):230–260, 2011.

Warszawski, Abraham. Multi-dimensional location problems. *Journal of the Operational Research Society*, 24(2):165–179, 1973.

Watts, Matthew E; Ball, Ian R; Stewart, Romola S; Klein, Carissa J; Wilson, Kerrie; Steinback, Charles; Lourival, Reinaldo; Kircher, Lindsay, and Possingham, Hugh P. Marxan with Zones: software for optimal conservation based land-and sea-use zoning. *Environmental Modelling & Software*, 24(12):1513–1521, 2009.

Weber, A. Über Den Standort Der Industrien, 1. Teil: Reine Theorie Des Standortes. (on the Location of Industries), 1909.

Weersink, Alfons; Clark, Steve; Turvey, Calum G, and Sarker, Rakhal. The effect of agricultural policy on farmland values. *Land Economics*, pages 425–439, 1999.

Weimer, David L and Vining, Aidan R. *Policy Analysis: Concepts and Practice*. Prentice Hall, 2005.

Weiszfeld, Endre. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, First Series*, 43(355-386):2, 1937.

Wheaton, William C. Monocentric models of urban land use: Contributions and criticisms. *Current Issues in Urban Economics*, pages 107–129, 1979.

White, Roger. Modeling multi-scale processes in a Cellular Automata framework. In *Complex Artificial Environments*, pages 165–177. Springer, 2006.

White, Roger and Engelen, Guy. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and planning A*, 25(8):1175–1199, 1993.

White, Roger and Engelen, Guy. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24(5):383–400, 2000.

White, Roger; Engelen, Guy, and Uljee, Inge. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B*, 24:323–344, 1997.

Whitley, Darrell and Kauth, Joan. *GENITOR: A Different Genetic Algorithm*. Colorado State University, Department of Computer Science, 1988.

Whitley, L Darrell. The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best. In *Proceedings of the Third International Conference on Genetic Algorithms (ICGA)*, volume 89, pages 116–123, 1989.

Wilkinson, Angela and Eidinow, Esther. Evolving practices in environmental scenarios: a new scenario typology. *Environmental Research Letters*, 3(4):045017, 2008.

Williams, John and Lawson, Rob. Community issues and resident opinions of tourism. *Annals of Tourism Research*, 28(2):269–290, 2001.

Wilson, Alan G. Theoretical geography: some speculations. *Transactions of the Institute of British Geographers*, pages 31–44, 1972.

Wilson, Alan G and Kirkby, Michael J. *Mathematics for Geographers and Planners*. Clarendon Press, 1975.

Wise, Steve. Green Infrastructure Rising. *Planning*, 74(8):14–19, 2008.

Wolfram, Stephen. Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, 10(1):1–35, 1984.

Wolfram, Stephen. *Cellular Automata and Complexity: Collected Papers*, volume 1. Addison-Wesley Reading, 1994.

Wong, David WS and Fotheringham, A Stewart. Urban systems as examples of bounded chaos: Exploring the relationship between fractal dimension, rank-size, and rural-to-urban migration. *Geografiska Annaler. Series B. Human Geography*, pages 89–99, 1990.

Wong, Koon Kwai. Urban park visiting habits and leisure activities of residents in Hong Kong, China. *Managing Leisure*, 14(2):125–140, 2009.

Woolley, Helen. *Urban Open Spaces*. Taylor & Francis, 2003.

World Health Organization (WHO), . Urban population growth, global health observatory database. `http://apps.who.int/gho/data/node.main.POP107?lang=en`, (Accessed 20 February 2014) 2013.

Wright, Jeffrey; Revelle, Charles, and Cohon, Jared. A multiobjective integer programming model for the land acquisition problem. *Regional Science and Urban Economics*, 13(1):31–53, 1983.

Wu, Fulong. A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computers, Environment and Urban Systems*, 20(6):367–387, 1996.

Wu, Fulong. SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 12(1):63–82, 1998.

Wu, Fulong. A parameterised urban cellular model combining spontaneous and self-organising growth. *Geocomputation: Innovation in GIS*, 7:73–85, 2000.

Wu, Fulong and Webster, Chris J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design*, 25(1):103–126, 1998. URL `http://econpapers.repec.org/RePEc:pio:envirb:v:25:y:1998:i:1:p:103-126`.

Wu, Fulong and Webster, Christopher J. Simulating artificial cities in a GIS environment: urban growth under alternative regulation regimes. *International Journal of Geographical Information Science*, 14(7):625–648, 2000.

Wu, Jianfeng; Zheng, Chunmiao; Chien, Calvin C., and Zheng, Li. A comparative study of Monte Carlo simple genetic algorithm and noisy genetic algorithm for cost-effective sampling network design under uncertainty. *Advances in Water Resources*, 29(6):899–911, 2006. ISSN 0309-1708. doi:10.1016/j.advwatres.2005.08.005. URL `http://www.sciencedirect.com/science/article/pii/S0309170805002071`.

Wu, JunJie and Plantinga, Andrew J. The influence of public open space on urban spatial structure. *Environmental Economics and Management*, 46(2):288–309, 2003.

Wu, Xiaoqing; Hu, Yuanman; He, Hong S; Bu, Rencang; Onsted, Jeff, and Xi, Fengming. Performance evaluation of the SLEUTH model in the Shenyang metropolitan area of northeastern China. *Environmental Modeling & Assessment*, 14(2):221–230, 2009.

Xiao, Ningchuan; Bennett, David A., and Armstrong, Marc P. Interactive evolutionary approaches to multiobjective spatial decision making: A synthetic review. *Computers, Environment and Urban Systems*, 31(3):232–252, 2007. ISSN 0198-9715. doi:DOI: 10.1016/j.compenvurbsys.2006.08.001. URL `http:`

//www.sciencedirect.com/science/article/pii/S0198971506000664. GeoComputation 2005.

Xie, Yichun. A generalized model for cellular urban dynamics. *Geographical Analysis*, 28(4):350–373, 1996.

Xie, Yichun; Batty, Michael, and Zhao, Kang. Simulating emergent urban form using agent-based modeling: Desakota in the Suzhou-Wuxian region in China. *Annals of the Association of American Geographers*, 97(3):477–495, 2007.

Xiujuan, Lei and ZhongKe, Shi. Overview of multi-objective optimization methods. *Systems Engineering and Electronics*, 15(2):142–146, 2004.

Xu, Xibao; Zhang, Jianming; Li, Jijun; Yu, Jianhua, and Zhou, Tong. Modeling urban land use changes in Lanzhou, China with GIS and DUEM. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Geoinformatics: Remotely Sensed Data and Information*, volume 6752, page 37. International Society for Optics and Photonics, 2007.

Xue, Shan; Fern, Alan, and Sheldon, Daniel. Scheduling Conservation Designs via Network Cascade Optimization. In *Twenty-Sixth Conference on Artificial Intelligence (AAAI-12), Toronto, Ontario, Canada*. Citeseer, 2012.

Yang, Lili; Jones, Bryan F, and Yang, Shuang-Hua. A fuzzy multi-objective programming for optimization of fire station locations through genetic algorithms. *European Journal of Operational Research*, 181(2):903–915, 2007.

Yang, Shuyuan; Wang, Min, and Jiao, Licheng. A quantum particle swarm optimization. *Proceedings of the 2004 Congress on Evolutionary Computation. CEC2004*, 16:320–324, June 2004.

Yeh, Anthony Gar-On and Li, Xia. A cellular automata model to simulate development density for urban planning. *Environment and Planning B*, 29(3):431–450, 2002.

Yeh, KC and Kwan, KC. A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. *Journal of Pharmacokinetics and Pharmacodynamics*, 6(1):79–98, 1978.

Yeniay, Ozgur. Penalty function methods for constrained optimization with genetic algorithms. *Mathematical and Computational Applications*, 10(1):45–56, 2005.

Zellner, Moira L. Generating policies for sustainable water use in complex scenarios: an agent-based model of Monroe County, Michigan. *Environment and Planning B: Planning and Design*, 34(4):664–686, 2007.

Zhang, Weixiong. Modeling and solving a resource allocation problem with soft constraint techniques. *Technical Report: WUCS-2002-13*, 2002.

Zhou, Xiaolu and Rana, Md Masud Parves. Social benefits of urban green space: A conceptual framework of valuation and accessibility measurements. *Management of Environmental Quality: An International Journal*, 23(2):173–189, 2012.

Zitzler, Eckart and Thiele, Lothar. Multiobjective optimization using evolutionary algorithms- A comparative case study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301. Springer, 1998.

Zitzler, Eckart and Thiele, Lothar. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.

Zitzler, Eckart; Deb, Kalyanmoy, and Thiele, Lothar. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2): 173–195, 2000.

Zitzler, Eckart; Laumanns, Marco, and Thiele, Lothar. SPEA2: Improving the strength pareto evolutionary algorithm. *Technical Report: TIK-Report 103*, pages 1–21, 2001.

Zucca, Antonella; Sharifi, Ali M, and Fabbri, Andrea G. Application of spatial multicriteria analysis to site selection for a local park: A case study in the Bergamo Province, Italy. *Journal of Environmental Management*, 88(4):752–769, 2008.